Modeling lung cell development using human pluripotent stem cells

Shuk Yee Ngan[1,*], Henry Quach[1,2,*], Joshua Dierolf[1], Jin-A Lee[1], Elena Huang[1,2], Amy P. Wong[1,2]

1. Program in Developmental and Stem Cell Biology, Hospital for Sick Children, 686 Bay Street, PGCRL 16-9420, Toronto, Ontario M5G0A4
2. Department of Laboratory Medicine & Pathobiology, University of Toronto

*Authors contributed equally to the paper

Keywords: Pluripotent stem cells, fetal lung, airway epithelia, organoids, air liquid interface, *in-vitro* differentiation, disease modeling, regeneration, development

Corresponding author:

Dr. Amy Wong, Program in Developmental and Stem Cell Biology, Hospital for Sick Children, 686 Bay Street, PGCRL 17-9704 Toronto, Ontario M5G0A4

Tel: 416-813-7654 x309334 | Email: apwong@sickkids.ca

ABSTRACT

Human PSC (hPSC) differentiations can capture developmental phenotypes and processes and are useful for studying fundamental biological mechanisms driving tissue morphogenesis and cell lineage development. Here, we show for the first time the temporal development of lung cell lineages using hPSC that model developmental milestones observed in primary tissue, the generation of renewable fetal lung epithelial organoids, and the functional utility of the lung models at different differentiation stages for Cystic fibrosis disease modeling. We first show the presence of hPSC-derived lung progenitor cells reminiscent of early trimester lung development and can capture a population enriched with basal stem cells that generates renewable airway organoids. Maturation and polarization in air liquid interface (ALI) generates additional epithelial cell lineages found in adult lung tissues including pulmonary neuroendocrine, brush, mature basal, ciliated and secretory cell types. Finally, pseudotime analysis of the integrated datasets from the fetal and ALI stages reveal the developmental trajectories of the cells as they emerge during differentiation. Overall, hPSC differentiation can capture aspects of human lung development and potentially provide important insight into congenital causes of diseases.

INTRODUCTION

Human lung organogenesis begins around 4 weeks of gestation as a small batch of cells emerging from the foregut endoderm. Co-development of the lung epithelium and the surrounding mesenchyme is facilitated by mutual cross-talk between these two cell types which ultimately guides the formation of an intricate airway tree[1,2]. In the first fetal stage of lung development, the pseudoglandular stage (GW5-17), formation of the major airways, respiratory parenchyma and other non-epithelial cell types contributes to the developing organ. Proximal-distal patterning during this stage leads to the first signs of epithelial lineage differentiation to form basal, ciliated, neuroendocrine (NE) and secretory cells. As the lung develops in the canalicular stage (GW16-27), additional cellular differentiation occurs with the formation of the distal lung and emergence of the primitive alveolar cells. The saccular stage (GW24-38) marks the expansion of air sacs and the production of surfactants in preparation for the lungs to switch to an air breathing organ after birth.[3] The final alveolar stage (GW24 up to 10 years) begins shortly before birth and continues well into childhood years with maturation of the alveoli and expansion of the capillaries, nerve and gas exchange regions of the lungs.

Single cell RNA sequencing has enabled our recent understanding of the complexities of the cellular phenotypes that make up the human and mouse lungs.[4–9] Importantly, the identification of new cell types or cell states have been implicated in human lung diseases[5,7,10] and the question of the origins of these cell types have sparked renewed interest in the understanding of human lung development[11,12]. New *in-vitro* model systems from isolated human fetal lung tissues or hPSC-derived are shedding light into key pathways regulating cell lineage formation through experimental manipulation[11,13–15]. Pluripotent stem cells are derived from the inner cell mass of the pre-implantation embryo (called embryonic stem cells, ESC), or are artificially generated

through reprogramming of somatic cells called induced pluripotent stem cells (iPSCs)[16,17]. Early step-wise differentiation of iPSC protocols modeling developmental milestones based largely on mouse lung development showed that bona fide lung epithelial tissues can be generated[18–20] and used to model cystic fibrosis airway disease[18,21–23]. Modifications in differentiation protocols and identifications of cell surface molecules for fluorescently activated cell sorting (FACS) of putative progenitor cells have significantly enriched for cells capable of forming 3-dimensional lung epithelial organoids for disease modeling[13,14,24–26] and model fetal development[11,27].

Human PSC-derived lung models are also currently paving the way for the identification of new drugs to treat individual cystic fibrosis (CF) disease[13,18,22,28]. Indeed iPSC derived from CF individuals has been shown to be a valuable asset for assessing new emerging CFTR modulators (small molecules aimed at targeting the genetic defect) at a personalized level[23,29]. We previously showed a directed differentiation protocol to generate robust CFTR expressing cells from human PSC and importantly CF iPSC-derived airway epithelia could be used to model CF disease *in vitro*[18]. Since then, others have demonstrated the use of CF iPSC-derived 3D airway organoids in forskolin-induced swelling (FIS) assays in the presence of CFTR modulator compounds as a method to test the efficacy of the compounds at a patient-specific level[14], similar to primary rectal organoids[30]. However, the long process of generating enough lung tissues from iPSC for rapid and high content screens is cumbersome and costly. The ability to generate another source of cells that expresses functional CFTR for rapid testing of CFTR modulator effectiveness can be beneficial.

Here we show an improved hPSC differentiation method to generate airway cells that model human lung cell development. We employed single cell RNA sequencing (scRNA-seq) to determine the cellular heterogeneity and the developmental course of the cells *in-vitro*. We validate the phenotypic and functional characteristics of the hPSC-derived lung cells throughout the

differentiation stages and correlate the expression patterns of hPSC-derived fetal lung (FL) cells to human pseudoglandular fetal lung tissues. We also show that FL cells can form 3D expandable epithelial spheroids containing multipotent basal stem cells and undergo multilineage differentiation in defined media or air liquid interface (ALI) culture. Finally, we show a developmental connection between cells that emerge during the differentiation process and highlight unique cell lineage relationships from the fetal to mature cell states. This may provide important clues for future cell engineering platforms for tissue regeneration and identify molecular regulations driving airway cell composition during homeostasis, development and disease.

RESULTS

Stage-specific characterization of *in-vitro* hPSC-derived fetal lung cells

We previously demonstrated a directed differentiation method to generate proximal airway epithelia expressing functional CFTR for CF disease modeling[18,22]. Integrating new developments in directed differentiation protocols toward lung cell lineages[11,15], we have refined our previous protocol to generate robust airway epithelial cell lineages (**Fig. 1**) mimicking a developmental trajectory of the cells as they differentiate and mature *in-vitro*. To demonstrate the efficiency of the new protocol, we directly compare both old (*Wong 2015*) and new protocols in subsequent analyses using two human pluripotent stem cell lines (human embryonic stem cell line CA1[18,22] and an iPSC lung reporter line NKX2-1^{GFP} [14,15]) previously used in development of differentiation protocols. Using a commercially defined kit to generate highly efficient definitive endoderm (>95%), cells are pushed towards anterior ventral foregut endoderm differentiation following a brief dual SMAD inhibition (transforming growth factor beta (TGFβ) and BMP inhibition) period (2 days) in media containing supplements favoring anterior foregut ventral endoderm (AFVE)

differentiation. At the end of this stage, >90% of the cells are SOX2$^+$ with no detectable midgut pancreatic and duodenal homeobox-1 (PDX1$^+$) and low level of hindgut caudal type homeobox-2 (CDX2$^+$) cells (4.2%) (**Supplementary Fig. 1A**). Gene expression analysis of foregut endoderm shows marked up-regulation of foregut markers forkhead box G1 (*FOXG1)*, *SOX2,* and modest elevation in the earliest lung marker *NKX2-1* (**Supplementary Fig. 1B**). Compared to control RNA (whole fetal lung from gestational week 18 tissue), airway epithelial specific lineage markers were not highly expressed by bulk gene expression analysis (using quantitative real-time reverse transcription polymerase chain reaction, qRT-PCR). Next, fetal lung (FL) epithelial cells from stage 3 of the differentiation were generated by adding Wnt pathway activator CHIR99021 and retinoic acid, both of which influences lung specification[31], to media supplemented with fibroblast growth factors (FGFs) and bone morphogenetic protein 4 (BMP4) favoring conditions that drive fetal lung growth[18]. Gene expression analysis by qRT-PCR of the cell cultures at this stage show abundant levels of *CFTR* expression. The high expression level of *CFTR* is reminiscent of first trimester fetal lung tissue where CFTR has previously been found to be abundantly expressed in the developing lung epithelium[32]. An increase in gene expression associated with basal (*TRP63*, transformation related protein 63 or TP63; and cytokeratin-5, *KRT5*), and ciliated (forkhead box J1, *FOXJ1*) cells were also observed (**Supplementary Fig. 2**).

To determine the heterogeneity of cell types and cell states in our *in-vitro* fetal lung cell culture, single cell RNA sequencing (scRNA-seq) was conducted on the FL cells (**Fig. 2A**). We first compared FL cells from our new differentiation to lung cells isolated from fetal lung tissue from two gestational week (GW) 16 and 18 lungs based on a pearson correlation coefficient (**Fig. 2B**). Our FL cells had high correlation with both GW 16 and 18 lungs (> 0.85) however, most clusters in our FL culture had higher correlation to our GW 16 lungs, with the exception of NKX2-

1^{hi}SOX9^{hi}SOX2^{lo}. Uniform manifold approximation and projection (UMAP) reveal mostly common cell types found in both new and old differentiation FL cultures with the exception of a cluster of NKX2-1^{neg}SOX9^{lo}SOX2^{hi} in our old protocol that lack *NKX2-1* expression but shares similar expression profile as NKX2-1^{hi}SOX9^{hi}SOX2^{lo} cluster (**Fig. 2C**). First, all "lung progenitor" cells expressed high levels of pan-epithelial genes *EPCAM*, and *CDH1* (E-cadherin) as well as *ID1* (Inhibitor of DNA binding 1, **Fig. 2D, Supplementary Fig. 3A**). Within the lung progenitor population, the clusters are subdivided by their expression of *NKX2-1, SOX9* and *SOX2*. The NKX2-1^{hi}SOX9^{lo}SOX2^{lo} lung progenitor cluster expressed elevated levels of sonic hedgehog signaling molecule (*SHH*), *SOX17* and the highest level of *CFTR*. The cluster of NKX2-1^{hi}SOX9^{hi}SOX2^{lo} lung progenitors expressed even higher levels of *ID1*, *ID2*, *ID3*, as well as *FOXP1*, Gata-binding protein-6 (*GATA6*) and *FOXA2* reminiscent of embryonic tip cells[31]. A cluster of NKX2-1^{hi}SOX9^{lo}SOX2^{hi} lung progenitors expressed high levels of cytokeratin-7 (*KRT7)*, insulin-like growth factor binding protein-5 (*IGFBP5*), and long non-coding RNA that harbours the gene miR-205 (*MIR205HG,* expressed in epithelial cells), reminiscent of embryonic stalk cells[31]. Specialized cell clusters including basal, ciliated, pulmonary neuroendocrine cells (PNEC), and mesenchymal cells were also found in both protocols[5,8,10]. Basal cells expressed the canonical marker *TP63* as well as *ID1, KRT19,* TNF Receptor Superfamily Member 12A (*TNFRSF12A*), and *MIR205HG*, while both cycling and non-cycling ciliated cells expressed *FOXJ1* and centromere protein U (*CENPU*)[33]. Cycling ciliated cells also expressed high levels of the proliferation gene DNA Topoisomerase II alpha (*TOP2A*) while ciliated cell precursors expressed high levels of *ID1*. Non-cycling ciliated cells expressed high levels of Forkhead Box N4 *(FOXN4*). Interestingly, *FOXN4* was also highly expressed in secretory precursors. Pulmonary neuroendocrine cells (PNEC) expressed the canonical gene Achaete-Scute Family BHLH

Transcription Factor 1 (*ASCL1*), as well as Chromogranin B (*CHGB*), Delta Like Canonical Notch Ligand 3 (*DLL3*), and Myocardial Infarction Associated Transcript (*MIAT*) while secretory precursors expressed *SCGB3A2, ID1,* Marker of Proliferation Ki-67 (*MKI67*), *TOP2A,* ADAM Metallopeptidase Domain 28 (*ADAM28*), and *MIR205HG*. Within the mesenchymal clusters, PDGFRB^{hi} cells expressed the highest level of platelet-derived growth factor receptor-b (*PDGFRB)*, as well as collagen alpha-1 (*COL1A1*), and *TOP2A* suggesting these cells are cycling mesenchymal cells. On the contrary, PLAT^{hi} mesenchymal cells expressed high levels of Plasminogen Activator, Tissue Type (*PLAT*) and lower *COL1A1*.

Feature map of the genes showed that *SOX2* is predominantly expressed in NKX2-1^{hi}SOX9^{lo}SOX2^{hi} progenitor cell, ciliated cell populations (ciliated cells, ciliated precursor, cycling ciliated cells) and PNEC suggesting proximal lineages while *SOX9* is predominantly expressed in NKX2-1^{hi}SOX9^{hi}SOX2^{lo} and sporadically in PDGFRB^{hi} mesenchymal cells and NKX2-1^{hi}SOX9^{lo}SOX2^{hi} clusters suggesting distal lineages (**Fig. 2E**). *ID1* is highly expressed in ciliated precursors and all lung progenitor cells. *FOXJ1* and *TP63* are expressed in ciliated cell and basal cell clusters respectively. We also found *CFTR* expressing cells predominantly in the lung progenitor populations of NKX-2-1^{hi}SOX9^{hi}SOX2^{lo} (10.9%), NKX-2-1^{hi}SOX9^{lo}SOX2^{lo} (24.1%), NKX2-1^{hi}SOX9^{lo}SOX2^{hi} (16.1%), and ciliated cells (16.7%) (**Fig. 2F**).

Next, we assessed the Gene Ontology (GO) terms associated with each cluster based on the differentially expressed genes (DEG) in both protocols and found processes related to cell signaling, proliferation, adhesion, motility, transcription and developmental processes (**Fig. 2G** and **Supplementary Fig. 3B**). Among the top 5 most significant GO terms in each cluster, we found all epithelial cell clusters (all lung progenitors, basal cells, all ciliated cells, secretory precursors and PNEC) and PLAT^{hi} mesenchymal cells expressed genes related to biological

pathways related to protein synthesis (e.g. regulation of transcription from RNA polymerase II promoter, regulation of mRNA stability, translational initiation, protein polyubiquitination). Additionally, GO terms related to cell division (cell division, mitotic nuclear division, sister chromatid cohesion) were significant in basal cells, cycling ciliated cells, ciliated cells and secretory precursors supporting the proliferative capacities of these cell types during the FL stage. Interestingly, NKX-2-1^{hi}SOX9^{lo}SOX2^{lo}, NKX-2-1^{hi}SOX9^{hi}SOX2^{lo} and NKX2-1^{hi}SOX9^{lo}SOX2^{hi} lung progenitors contains uniquely significant GO terms associated with epithelial cell differentiation, establishment or maintenance of epithelial cell apical/basal polarity, and muscle-interacting biological pathways respectively. The unique GO terms among the three progenitor clusters resembles the different roles of lung progenitors during branching morphogenesis in late pseudoglandular stage[31]. PNEC is significantly associated with multiple neurological-related pathways (e.g. sympathetic nervous system development, neuron migration) which supports the neurological sensory function of PNEC[34]. Ciliated cells were associated with GO terms related to positive regulation of transcription from RNA polymerase II promoter and nervous system development, and cycling ciliated cells and ciliated precursors were associated with biological processes related to cell proliferation and transcription-related pathways respectively. Importantly, ciliated precursors were significantly associated to cilium-related processes such as cilium assembly and cilium morphogenesis. Basal cells and ciliated precursors also expressed genes associated with Wnt signaling pathways (e.g. Wnt signaling pathway, planar cell polarity pathway, positive/negative regulation of canonical Wnt signaling pathway) which were previously shown to regulate ciliated cell differentiation[35]. PLAT^{hi} mesenchymal cell GO analysis revealed processes related to protein synthesis pathways, and non-lung differentiation pathways. GO analysis of PDGFRB^{hi} mesenchymal cells, on the other hand, showed processes related to extracellular matrix,

bone and muscle-related biological pathways. The distinctive GO terms between PLAT^{hi} and PDGFRB^{hi} mesenchymal cells suggest different phenotype and functional roles of these cells in the FL cultures.

Immunofluorescence staining of FL cultures for multipotent progenitor cell markers (SOX2, SOX9) and lung marker NKX2-1 (or thyroid transcriptional factor 1, TTF1, protein *herein*) showed TTF1^{+} cells expressing both SOX2^{+} and SOX9^{+} as well as TTF1^{+}SOX2^{+} and TTF1^{+}SOX9^{+} double positive cells were found in both protocols (**Fig. 3A** and **Supplementary Fig. 3C**) using both hPSC lines, suggesting an emergence of a heterogeneous lung progenitor population in the FL cultures. Using fluorescence activated cell sorting (FACS) analysis, a general increase in TTF1+ cells and TTF1^{+}SOX2^{+} and TTF1^{+}SOX9^{+} double positive cells were observed (**Fig. 3B**). This is in line with findings that the developing human distal tip lung contains multipotent progenitors that express both SOX2 and SOX9[33] and with previous differentiation studies demonstrating fetal-like lung progenitors after differentiating anterior ventral foregut endoderm cells[14,27].

Using CD47 as a marker of "lung progenitor" as previously described[36], a significant increase in CD47^{+}TTF1^{+} double positive cells were found in the new FL cultures. Moreover, FL cells in the new differentiation expressed abundantly more CFTR localized to the plasma membrane than our previous protocol (**Fig. 3A**). Isotype controls using non-immune immunoglobulins were used to assess non-specific binding and background fluorescence for imaging.

Recent studies have identified fetal basal stem cells at the FL stage of differentiation as marked by expression of the *TP63*[11,13]. We also show improved generation of P63^{+} basal stem cells (>16% in CA1 and >51% in BU3-NKX2-1^{GFP}) in our new protocol, not surprisingly, with

differences in efficiency in two independent PSC lines (**Fig. 3B**). Indeed, increased expression levels of genes associated with basal cells (*TRP63* and *KRT5*) as well as ciliated cells (*FOXJ1*) was observed (**Supplementary Fig. 2**).

Importantly, higher protein levels of the mature CFTR protein Band C (representing a fully glycosylated functional form of the protein expressed on the plasma membrane, ~170 kDa) was observed in our new differentiation protocol compared to our previous protocol (**Supplementary Fig. 3D**). Using our previously established modified apical chloride conductance (ACC) assay measuring membrane potential changes through changes in fluorescence intensity (Fluorescence imaging plate reader, FLiPR,[23] we found CFTR activity was improved from old protocol (up to 20% CFTR function, **Fig. 3C**) to the new protocol (40% CFTR function, **Fig. 3D)** upon forskolin-induced CFTR activity. Specificity of CFTR function was confirmed using CFTR inhibitor-172 treatment. Both hPSC lines demonstrated enhanced CFTR function averaging around 32% function in our new protocol compared to our old (~10% function) (**Fig. 3E**) and thus provide a powerful source of cells for rapid high throughput functional screens of novel CFTR modulators[37].

Generation of renewable 3D lung organoids

To determine whether isolated FL cells embedded in matrigel could generate lung epithelial organoids (spheroids) (**Fig. 4A**), we sorted for pan-epithelial cell adhesion molecule (EPCAM or CD326) and found the majority of the FL cells were $CD326^+$ (up to 72%, **Fig. 4B**). Within 4 days, epithelial organoids emerged (**Fig. 4C**). These organoids were capable of propagation in expansion media (**Fig. 4D**) containing dual smad inhibitors for up to 10-15 passages with decreasing expansion potential in late passages (data not shown). We tested this organoid generation potential in three independent hPSC lines and found varying efficiencies of organoid generation (**Fig. 4E,**

F) from CD326^{+} single cells. Renewable epithelial organoids are composed of mostly ΔNP63^{+} basal stem cells (>90%, **Fig. 4G, H**), confirmed using a double fluorescence reporter hPSC line BU3-NKX2-1^{GFP}/TP63^{tdTomato} (NGPT) where tdTomato fluorescence was observed making up the majority of the cells in the organoids. The abundance of basal stem cells suggests the expansion capacity of the organoids may be driven by these multipotent progenitor cells as previously described[13]. The organoids can be cryopreserved and expanded without losing cell phenotype after serial passaging (data not shown). Phenotypic characterization of the self-renewing organoids shows cells expressing TTF1, SOX2, SOX9 and luminal expression of ZO-1. CFTR appears localized to the cytoplasm of the cells (**Fig. 4I**). Notably, the majority of these cells express ΔNP63 and cytokeratin-14 (KRT14), associated with basement membrane localized basal cells in the airways.

To determine the multi-epithelial differentiation potential of the basal stem cells in the epithelial organoids, epithelial organoids were cultured in differentiation media for up to 14 days (**Fig. 5A**). Morphologically, hPSC-derived epithelial organoids share morphological similarities to organoids derived from primary bronchial epithelial cells (**Fig. 5B**). Characterization of the differentiated organoids show up-regulation of several genes associated with specific epithelial cell lineages found in the airways including basal cell marker *KRT5*, ciliated cell marker *FOXJ1*, luminal epithelial cell marker *KRT8*, and secretory cell marker *MUC16* (**Fig. 5C**) and is confirmed with the presence of ΔNP63 and KRT14, acetylated a-tubulin, cytokeratin-8/18 (KRT8/18), and secretoglobulin 1A1 (SCGB1A1) and mucin 5ac (MUC5AC) respectively in bulk qPCR analysis (**Fig. 5D**). A near two-fold up-regulation of *CFTR* is also observed and can be found on the apical membrane (lumen side) of the organoid along tight junction protein ZO-1 (**Fig. 5D**). The lung

(TTF1$^+$) epithelial (E-cadherin$^+$, panCK$^+$) organoids express proximal marker SOX2 suggesting these are airway epithelial organoids and do not contain cells of the distal respiratory phenotype.

To assess the functional property of the airway epithelial organoids, we assayed for CFTR protein expression for the presence of Band C (~170 kDa) indicative of a mature apically. localized protein capable of function. Only organoids derived from CD326$^+$ cells express mature CFTR (**Fig. 5E**). To validate this, we performed FLiPR analysis on organoids 2-3 days upon seeding onto transwells and found an increase in fluorescence activity indicative of CFTR function upon forskolin-induction that was inhibited with CFTR inhibitor-172 (**Fig. 5F**). Assessing the CFTR functional potential of several organoids from different sets of differentiations show on average up to 10% CFTR function can be detected in the differentiated organoids (**Fig. 5G**). Importantly, a decrease in CFTR function in differentiated cells aligns with a decrease in CFTR expression in fetal lung tissues during development beyond the second trimester where CFTR is found only in a subset of bronchial epithelial cells[32,38].

Generation of mature adult-like airway epithelial tissue

Next, we assessed the ability to generate functional mature airway epithelial tissue by culturing the cells on collagen-coated transwells as previously described[18] to generate air liquid interface (ALI) cultures (**Fig. 6A**). During this period, the epithelial cells undergo additional differentiation, maturation and polarization to form a pseudostratified epithelium reminiscent of the mature airways. Our modified ALI protocol entails a two-step method that first allows expansion of the cells seeded onto the transwell membrane to form a monolayer of cells prior to differentiation and polarization. Characterization of the genes expressed in the cell population in the ALI cultures show an average significant up-regulation of *CFTR* expression in the new protocol by >500-fold

compared to the old protocol while many of the other epithelial cell lineage markers remained relatively consistent (**Supplementary Fig. 4)**. Importantly, there were differences in expression levels for many of the genes between the two hPSC lines tested.

For higher resolution of the cell heterogeneity in our cultures, we performed scRNAseq on the ALI epithelium after 5 weeks in culture. UMAP analysis shows presence of additional cell types observed in the adult airways including Goblet cells, secretory club cells, ionocytes and Brush/PNEC cell types as well as cycling basal and ciliated cells (**Fig. 6B**), the latter two expressing high levels of proliferation markers *MKI67* and *TOP2A* (**Fig. 6C** and **Supplementary Fig. 5A**)[5,8,10]. Basal cells express high levels of *TP63, KRT5, KRT14, KRT19,* and *KRT7* as previously described[5,11]. Cycling basal cells express less *TP63* but higher Sclerostin Domain Containing 1 (*SOSTDC1*), previously found in fetal basal cells with multipotent differentiation capacity[9]. Both cycling ciliated cells and ciliated cells express canonical marker *FOXJ1* however, CFTR is predominantly expressed in the non-cycling ciliated cell population. Goblet cells expressed high levels of anterior gradient protein homologs *AGR2, AGR3,* lysozyme *LYZ,* trefoil factors *TFF1*, *TFF2, TFF3*. Secretory precursors express less secretory proteins and higher levels of *MKI67* and *TOP2A*. Secretory club cells expressed high levels of cytokeratins *KRT4, KRT19,* mucin *MUC16*, and secretoglobulin *SCGB3A2* while Brush/PNEC expressed insulin growth factor *IGF1*, sclerostin domain containing protein *SOSTDC1*, and Neuron derived neurotrophic factor *NDNF*. A cluster of unknown cells expressing high levels of *IQGAP2, ID3, ID4* suggesting these may be pre-ionocyte[8] cells, although without expression of the canonical genes *FOXI1* and *CFTR,* the phenotype remains unresolved. Interestingly, *CFTR* was expressed in most epithelial cell types including goblet, secretory club, ciliated and basal cells and unlike previous findings[8], to a lesser extent ionocytes. Epithelial cells undergoing epithelial-mesenchymal transition (EMT) were also

found in the ALI cultures and expressed both epithelial cytokeratins (*KRT19, KRT7*) and canonical mesenchymal cell genes platelet-derived growth factor receptors *PDGFRA*, and *PDGFRB*, suggesting the need for supporting mesenchymal cells in the maturation of the airways[8,39]. Our new differentiation protocol also identified a cluster of POSTN2^{hi} mesenchymal cells that express common genes found in the myofibroblast cluster with the exception of a slightly higher expression of Periostin (*POSTN2*), previously found in human lungs that secrete extracellular matrices such as collagen[40]. Unlike previously found, *CFTR* expression is highly expressed within the ciliated cell cluster (**Fig. 6D**). While an unknown cluster of cells that may represent pre-ionocyte cells exist, these cells do not express *CFTR*. Assessing the number of CFTR-expressing cells in the ALI culture shows the majority of these cells are ciliated followed by goblet and secretory club cells (**Fig. 6E**). The reduced expression of *CFTR* in clusters associated with EMT and myofibroblasts implicates the role of CFTR in epithelial-mesenchymal transition[2].

Gene ontology analysis of each cluster showed the top 5 most significant GO terms generally related to cell cycle regulation, cell matrix organizations, immune pathways and developmental signaling pathways for the ALI cultures (**Fig. 6F** and **Supplementary Fig. 5B**). We found all clusters in both protocols contained GO terms related to multiple cell adhesion/matrices-related biology pathways (cell-cell adhesion, adherens junction organization, cell adhesion, bicellular tight junction assembly, collagen catabolic process) which resembles the active transition of extracellular matrices in the lung[41]. Cell division, mitotic nuclear division, sister chromatid cohesion, and G1/S transition of mitotic cell cycle were found in cycling basal cells, cycling ciliated cells and Brush/PNEC suggesting the proliferative nature of these cell types. We also found immune response related GO terms in secretory and basal cells which supports their role in regulating immunity in the epithelium[42]. Ciliated cell was associated with the sphingolipid

biosynthesis process and is important in regulating cilia functions and formations[43]. The unknown cluster interesting had GO terms associated with cell proliferation (positive/negative regulations of proliferation), extracellular matrix (extracellular matrix organization, collagen fibril organization) and angiogenesis-related (angiogenesis, blood vessel development, sprouting angiogenesis) biological pathways.

Immunostaining of the mature ALI cultures in our new protocol shows TTF1 expression is restricted to some cell types, and the majority co-expressed with SOX2 indicates proximal airway phenotype (**Fig. 7A** and **7A'**) similar to adult lung tissue (**Fig. 7A"**). A pseudostratified airway epithelium composed of basolateral-localized basal cells (ΔNP63^{+}) and lumenal secretory (SCGB3A2^{+}) and ciliated (acetylated alpha-tubulin^{+}) cells are extended to the luminal/apical side (**Fig. 7B**) similar to adult airway tissue (**Fig. 7B'**). Non-immune reactive immunoglobulins (IgGs) were used as controls for non-specific binding (**Fig. 7C**). Bulk gene expression analysis shows both hPSC lines are capable of generating many of the mature airway epithelial cell types observed in airways using the new protocol (**Supplementary Fig. 4**). Protein analysis of CFTR shows high levels of Band C CFTR protein species (**Fig. 7D**). Importantly, a near 3-fold higher CFTR function can be detected in the new ALI cultures compared to our previous protocol (**Fig. 7E, F**).

Cell lineage relationship of differentiated cells

Finally, to determine the lineage relationships of the cells as they differentiate from fetal to ALI stages, we performed pseudotime analysis of the integrated FL and ALI clusters using Monocle3[45]. To do this, we combined the data sets of FL and ALI cultures using Seurat V3's built-in integration algorithm[44] which identified anchor genes between the two data sets that were used to subsequently merge the datasets. After integration, we first generated an integrated UMAP outlining the cell

clusters (**Fig. 8A**). Pseudotime analysis was then performed to infer the cell lineage trajectories and the temporal relationship of the cells based on gene expression changes through pseudotime. Using the built-in partition-based graph abstraction algorithm in Monocle3[45], the integrated clusters were partitioned into 5 groups ('parts') (**Fig. 8B**), and within each 'parts' cell trajectories were identified (**Fig. 8C**). We then inferred the lineage relationships between the cells based on their differentiation state and pseudotime trajectory (**Fig. 8D**). Part 1 and 4 appeared to share similar lineage trajectories. First, a relationship between FL ciliated cells and Brush/PNEC and PLAT^{hi} mesenchymal cells was found. FL ciliated cells appeared to also derive various stromal PDGFRB^{hi} mesenchymal, myofibroblasts, and POSTN2^{hi} mesenchymal cells through EMT. In part 2, pseudotime analysis revealed two distinct lineage trajectories. First, NKX2-1^{hi}SOX9^{lo}SOX2^{hi} progenitors are the source of NKX2-1^{hi}SOX9^{hi}SOX2^{lo} cells and then NKX2-1^{hi}SOX9^{lo}SOX2^{lo} progenitors. Goblet cells in ALI are derived from secretory precursors stemming from NKX2-1^{hi}SOX9^{lo}SOX2^{lo} progenitors in the FL stage. Additionally, the secretory/ciliated trajectory showed FL basal cells to be directly and most closely linked to basal cells in ALI, and the source of FL secretory precursors that gave rise to secretory club cells and FL ciliated cells. Ciliated cells in ALI emerged from FL ciliated cells possibly through a cycling ciliated cell population. In part 3, FL ciliated cells appeared to emerge from ciliated cell precursors. Interestingly, the gene expression programs of the cycling ciliated cells in ALI resembles that of FL ciliated cells suggesting a "de-differentiation" program or reversion to premature state during cell proliferation. Cycling basal cells in ALI also shared similar gene expression programs as ciliated cell precursors which was not surprising as the latter expressed many genes previously identified in basal cell subclasses (data not shown). Finally in part 5, an obscure trajectory dominated by FL PNEC cells was found.

MATERIALS & METHODS

Embryonic stem cell (ES) and Induced pluripotent stem cell (iPSC) expansion and maintenance

All human stem cell lines were approved for use by the Canadian Stem Cell Oversight Committee SCOC and the SickKids Research Ethics Board (REB # 1000071246 for use of human ES cells). Human ES line CA1 was provided courtesy of Dr. Andras Nagy (Lunenfeld Tanenbaum Research Institute, Toronto) and iPSC lines BU3-NKX2-1^{GFP} and BU3-NKX2-1^{GFP}/TP63^{tdTomato} (NGPT) were provided by Dr. Darrell Kotton (CReM Boston University, Boston). The hPSC lines were cultured in serum-free defined conditioned media mTeSR or mTeSR plus (STEMCELL Technologies) on hES-certified matrigel (Corning cat#354277). Briefly, cell culture plates were pre-treated with diluted Matrigrel (1:25 dilution) for at least 1 hour at room temperature or overnight at 4 °C. Cells were grown to ~80% confluency before passaging using ReLeSR™ (STEMCELL Technologies, catalog #05872). Cells were split 1:6 with each passage. Media change was performed daily using mTeSR. Frozen vials of iPSC were thawed into mTeSR plus media supplemented with 10 μM ROCK inhibitor (Y27632, STEMCELL Technologies, catalog #72304) and the media changed after 2 days.

Modified differentiation of iPSC into lung models

The new differentiation method was adapted and modified from our previously published protocol[18,22]. Briefly, iPSCs are differentiated into definitive endoderm (Stage 1) using STEMCELL Technologies StemDiff DE kit (*Stage 1 media*), as per manufacturer's protocol (catalog #05110). Definitive endoderm cells are then differentiated into anterior ventral foregut endoderm (Stage 2) using dual SMAD inhibition for 2 days with 2 μM Dorsomorphin

(STEMCELL Technologies, cat#72102) and 10 μM SB431542 (STEMCELL Technologies, cat#72234) in media containing 500 ng/mL fibroblast growth factor-2 (FGF2, Preprotech, cat#100-18B) and 50 ng/mL recombinant sonic hedgehog (SHH, Preprotech, cat#100-45) in Knockout DMEM (Gibco cat #10829-018), 10% Knockout serum replacement (KOSR, Gibco cat# 10828-028), 1% penicillin/streptomycin, 2 mM Glutamax (Gibco, cat# 35050-061), 0.15 mM 1-mono-thioglycerol (MTG, Sigma-Aldrich, cat# M6145), and 1 mM non-essential amino acid (Gibco cat# 11140-050) (*Stage 2 media*). Cells were then switched to media containing only FGF2 and SHH for the remaining 3 days. On day 10 of differentiation (*Stage 3a media*), lung epithelial progenitors were induced using 10 ng/mL FGF7 (KGF, Preprotech, cat#100-19), 50mg/ml FGF10 (Preprotech, cat#100-26), bone morphometric protein-4 (BMP4, R&D Systems, cat#314-BP), 3 μM CHIR99021 (STEMCELL Technologies, cat#72054), 100 nM Retinoic acid (Sigma cat# R2625), and 10 μM Y27632 for the first 2 days and without Rock inhibitor thereafter in lung basal media composed of IMDM/F12, N2 supplement-A (STEMCELLTechnologies, cat#07152), B27 supplement without Vitamin A (Thermo Fisher Scientific, cat#17504044), 2 mM Glutamax, 0.15 mM MTG, 0.5% bovine serum albumin (BSA, Thermo Fisher Scientific, cat#12483020), 50 μg/mL ascorbic acid (Sigma-Aldrich, cat# A4544). On day 16 (*Stage 3b media*), proximalization of the lung progenitor cells was induced with media containing 50 ng/mL FGF7, 50 ng/mL FGF10, and 10 ng/mL FGF18 (Sigma-Aldrich, cat# F7301) in lung basal media for 5 days. On day 21 (*Stage 4 media*), media was switched to PneumaCult-Ex plus containing 1 μM forskolin, 10 ng/mL FGF18, 50 ng/mL FGF7, and 50 ng/mL FGF10 for an additional 5 days minimum. Cells can be cultured in this media for at least 10 days. To improve survival from the media switch, 10 μM Y27632 for the first 2 days was added to the media. On day 26 (*Stage 5a media*), stage 4 cells or airway organoids (AO), were seeded onto collagen-IV (Sigma Aldrich, catalog #C7521)-coated

transwells (Thermo Fisher, Corning cat#01-200-161) at a seeding density of $2.5x10^5$ cells/insert. Cells were expanded to form a monolayer of epithelial cells on the transwell membrane prior to differentiation in Stage 5a media containing 2 μM Dorsomorphin, 10 μM SB431542, and 10 μM DAPT (Abcam, cat# ab120633) in PneumaCult-ALI media (supplemented with 50 nM hydrocortisone, 50 nM dexamethasone, and 10 μM Y27632 for the first 2 days of culture). Cells were maintained in Stage 5a media on the apical side of the transwell until a monolayer of cells were formed. Once a confluent monolayer was formed, media (*Stage 5b media)* comprised of PneumaCult-ALI media supplemented with 2 μM Dorsomorphin, 10 μM SB431542, and 10 μM DAPT and added only to the basolateral side of the transwells to induce differentiation and polarization of the epithelium for up to 4 additional weeks.

Fetal lung cells at the end of stage 3 of the differentiation is sorted for pan-epithelial cell surface protein CD326 using magnetic activated cell sorting (MACS, Miltenyi) using LS columns (cat# 130-042-401). Non-epithelial (CD326-) cells were discarded and CD326+ single cells were seeded onto collagen IV-coated 96 well black/clear bottom plates (Thermo Fisher, cat# 165305) at a density of 25,000-50,000 cells/well. The following day, non-adherent cells were removed and replaced with fresh Stage 3b media. Once the cells were confluent (90%), CFTR functional studies by FLiPR were performed. CF iPSC-derived cultures were pre-treated with CFTR modulators prior to functional measurements.

To generate airway organoids (AO), epithelial cells at the end of Stage 3 or Stage 4 were isolated by MACS as above. Single cells ($CD326^+$) were resuspended in growth factor-reduced matrigel (BD354230) in 50 μL volume and immediately transferred to cell culture plates to form matrigel "domes". Matrigel was allowed to solidify at 37 °C for 20 minutes before media was added covering the dome for culture. For AO expansion of cells isolated from Stage 3, *Expansion*

Media comprised of Pneumacult Ex Plus (STEMCELL Technologies, cat# 05040) supplemented with A83-01 (STEMCELL Technologies, cat# 72022) and DMH-1 (STEMCELL Technologies, cat# 73632)[13]. For AO differentiation, *Expansion Media* was replaced with *Differentiation Media* comprised of Pneumacult-ALI complete media, 10 μM Y27632, 10 μM SB431542, 3 μM CHIR99021, 10 μM DAPT, 10 ng/mL FGF18, 10 ng/mL FGF7, 10 ng/mL FGF10, 1 μM forskolin, 1 μM IBMX, and 3 μM ATRA (Millipore cat # R2625) for at least 7 days. Media change occurred every other day. Within 7 days, AO can be observed with characteristic bilayer concentric rings as observed in organoids derived from human primary bronchial epithelial cells. For swelling assays, 5000 CD326+ cells were seeded in semi-solid matrigel (50% matrigel in organoid media) on the apical side of polyester transwells. The mixture was allowed to solidify at 37 °C for 45 minutes before the bottom chamber was filled with 1.5mL of organoid media. Media change occurred every other day.

The above protocol was compared to directed differentiation protocols from our previous method[22]. Key differentiation stages were directly compared using two human PSC lines (human ES line CA1 and iPSC line BU3-NKX2-1^{GFP}).

Quantitative Real-time PCR

RNA was extracted using the RNeasy Kit (Qiagen, cat#74106) as per manufacturer's protocol. Prior to RNA isolation, genomic DNA was removed using the QIAshredder (Qiagen, cat# 79656). Isolated RNA was reverse transcribed with Superscript II (Thermo Fisher Scientific, cat#18-064-071) as per manufacturer's protocol. Quantitative real-time PCR was performed with PowerUP SYBR Green Mastermix Master Mix on ViiA7 (Applied Biosystems). Gene expression is

normalized to house-keeping gene GAPDH and expressed relative to control human tissue (**Supplementary Table 1**) RNA extracts (2^ΔΔCT). A total run of 40 cycles. Cycle threshold (CT) values above 38 were considered "not expressed". Primer sequences are listed in **Supplementary Table 2**.

In vitro Single cell RNA-sequencing (scRNAseq) preparation and sequencing

Samples were dissociated into single cell suspension with viability over 80%. For lung progenitor cultures, cells were washed twice with PBS, and dissociated with trypsin (Wisent, cat# 325-043-CL) at 37 °C for 4 minutes, then inactivated by a sorting solution, PBS containing 1% FBS. For ALI cultures, media were removed from both apical and basal chambers, and washed twice with PBS. Trypsin was added in both the apical and basal chambers, and incubated at 37 °C for 5 minutes. Trypsin in the basal chamber were removed, and cell pellets were collected with a sorting solution. For airway organoids, Matrigel was degraded by dispase (STEMCELL Technologies, cat# 07923) at 37 °C for 1 hour. After 2 PBS washes, trypsin was added to dissociate organoids at 37 °C for 5-10 minutes, then inactivated by sorting solution. After centrifugation at 300g for 5 minutes, the cell pellets were resuspended in sorting solution, and strained by 0.4 μm strainers (Falcon, cat# 352340). MULTI-seq barcoding was performed as per the protocol provided by Dr. Zev J Gartner Lab[46]. Cell number and viability were assessed by trypan blue (Gibco; cat# 15250061) with Countess II (Life Technologies, cat#A27977). Four Multi-seq barcoded samples were combined as one sample, and loaded 25,000 cells per GEM beads. GEM generation & barcoding, post GEM-RT cleanup & cDNA amplification, and Library construction steps were performed as per Chromium Next GEM single Cell 3' Reagent Kits v3.1 (10x Genomics,

cat#1000121, 1000120, 1000123). Sequencing was performed on the NovaSeq6000 at a read depth of approximately 60 000 reads/cell. (TCAG Facility, SickKids).

Preparation of single cell suspension from primary fetal lung tissue

GW 16 and GW18 fetal lung samples were procured from the Research Centre for Women's and Infants' Health (RCWIH) Biobank (SickKids REB# 1000067499 and LTRI REB #20-0035-E) and placed immediately on cold 1XPBS. Single cell suspension was achieved using the 37_C Multi_tissue_dissociation B program on the gentleMACS octo-dissociator (Miltenyi Biotec). Subsequently, cells were gently strained with 0.4 µm strainers. Lastly, cells were resuspended in RBC lysis buffer (Invitrogen, cat# 00-4333-57) for two minutes and cell viability was determined as above. Single cell suspension was loaded at 6000 cells per GEM beads. GEM generation & barcoding, post GEM-RT cleanup & cDNA amplification, and Library construction steps were performed as per Chromium Next GEM single Cell 3' Reagent Kits v3.1 (10x Genomics, cat#1000121, 1000120, 1000123). Sequencing was performed on the NovaSeq6000 at a read depth of approximately 60 000 reads/cell. (TCAG Facility, SickKids).

Overview of single cell RNA-sequencing analysis pipeline

Single cell sequencing analyses were mainly performed with *Seurat v.4.0* as previously described[47]. Briefly, untrimmed reads were generated using *supernova/cellranger* mkfastq and bcl2fastq2 v2.20. Reads alignment to human reference genome and generation of gene expression matrix were done using 10x Genomics *Cell Ranger v6.0.1* software with provided hg19 (GRCh38) reference genome and transcriptome annotation. MULTI-seq barcodes were demultiplexed using *deMULTIplex*[46], an available R package provided by the Gartner Lab. In each sample, to ensure

high data quality for further analysis, cells with more than 20,000 or less than ~1,000 genes or mitochondrial transcript fraction over 10% were filtered out. Genes expressed in less than 3 cells were excluded from our analysis. Cell cycle phase was scored by expression levels of cell cycle related genes. Cellular variance of total number of UMI, mitochondrial transcript fraction, and cell cycle phase was regressed out and gene expression levels were normalized via *SCTransform*[49]. For the *in-vitro* cultures, similar lung differentiation cultures derived from different cell lines were merged into one dataset through the Seurat's *merge* function. In the merged samples, principal component analysis (PCA) was performed based on normalized (Pearson residuals) expression levels of genes. Graph-based (Louvain) clustering approach and UMAP dimension reduction was performed based on between the top 50 principle components. Primary fetal lung datasets were subsetted for cells expressing EPCAM for correlation analysis. Identification of positive markers for each cluster was performed with the two-sided Wilcoxon rank sum test[48]. Differential gene expression analysis was used to manually annotate based on previously published work.[5–11,13,33,49]

Each *in-vitro* differentiation cultures were analyzed separately. To quantify the transcriptome similarity between *in-vitro* cells and control fetal lung cell sub-clusters, Pearson's correlation coefficients of expression levels were calculated using shared genes of primary and *in-vitro* cells. Gene ontology enrichment in cluster markers was performed using *DAVID v6.8*[50,51]. Significant enrichment was defined as Benjamini-corrected $P < 0.05$. Pseudotime analysis of samples were performed using *Monocle3*[52]. In brief, datasets from fetal lung and air-liquid interface cultures were integrated using Seurat's integration pipeline. The resulting dataset was divided into partitions[45] and its trajectories were subsequently analyzed through Monocle3's *learngraph()* function.

Flow cytometry

Samples were dissociated with TrypLE for 5 minutes at 37 °C, and collected at 300g for 5 minutes. The cells were washed once with 3 ml of FACS buffer, and then transferred to a FACS tube containing a filter top to remove cell clumps. Single cell suspensions in 200 ul FACS staining buffer (PBS -Ca2+/Mg2+ with 0.5% BSA) were stained with Fixable Viability Dye eFluor450 (eBioscience, cat# 65-0863-14) for 30 minutes on ice. Cells were then washed 3 times prior to fixation with 4% PFA for 30 minutes at room temperature. Primary antibodies (**Supplementary Table 3**) were added to resuspended fixed cells in 200ul FACS staining buffer at 4 °C for 30 min followed by secondary antibody staining (if needed) at 4 °C for another 30 mins. After a brief wash, the cells were resuspended in FACS buffer and analyzed with the BD Biosciences LSRII analyzer. Data analysis was performed using the FlowJo software.

Immunofluorescence

Monolayer cultures were washed twice with PBS (without Ca^{2+} and Mg^{2+}) and then fixed with ice-cold methanol for 20 minutes, followed by another round of wash with PBS. Cells were subsequently permeabilized with 0.25% Triton X-100 for 20 minutes. The cells were then treated with a blocking solution containing 5% BSA and 0.25% Triton X-100 to block non-specific binding at room temperature for 30 minutes to 1 hour. Primary antibodies (**Supplementary Table 3**) were added to the blocking solution and incubated overnight at 4℃. Cells were then washed 2X with PBS followed by incubation with secondary antibodies (**Supplementary Table 3**) at room temperature for 2 hours. Finally, another wash was performed and the nuclei counterstained with DAPI for 15 minutes at room temperature. Stained cells were kept in PBS and images captured using the Olympus Spinning Disc Confocal within 2 days of staining.

Organoids were released from the matrigel bubble using ice cold dispase (STEMCELL Technologies, cat# 07923) and kept on ice for about 1 hour with periodic mixing to gently melt the matrigel. The organoid suspension was then centrifuged at 300 g for 3 minutes to pellet the organoids. After careful removal of the melted matrigel in the dissociation buffer, the organoids were resuspended in fixative (4% PFA) for 15 minutes at room temperature. After a brief wash with PBS containing 0.25% TritonX-100, the organoids were resuspended in PBS containing 5% BSA and 0.1% TritonX-100 and kept at 4 ℃ until staining. For organoid staining, organoids were incubated with primary antibodies for overnight incubation at 4 ℃. After two washes with PBS, secondary antibodies and Hoescht dye (1:1000, Thermo Fisher Scientific, cat# H3570) were added for 1-2 hour incubation at room temperature. The organoids were washed with PBS twice before mounting on superfrost slides (Fisher Scientific, cat# 22-037-246). 30ul droplets containing the organoids were gently placed onto the slide and pre-mounted with 0.12mm Secure-Seal silicone spacers (Thermo Fisher cat# S24737) before securing into place with a coverslip. Organoids were imaged within 2 days of staining and stored in a humidified chamber at 4 ℃ protected from light until then. Fluorescence images were captured with the Olympus spinning disc confocal microscope and analyzed with the Volocity imaging software (Quorum Technologies).

For fixed paraffin-embedded slide staining of ALI cultures, slides were first deparaffinized and antigen retrieval was performed using either citrate buffer or HIER solution (Abcam, ab208572). Successive steps were followed as previously described[22].

Air-liquid Interface Apical chloride conductance (ACC) Assay for CFTR function

The ACC assay was used to assess CFTR mediated changes in membrane depolarization using methods as previously described[23]. For ACC measurements, basal chamber were filled with HBSS (Wisent, cat# 311-513), and the cells on transwells were incubated with zero sodium, chloride and bicarbonate buffer (NMDG 150 mM, Gluconic acid lactone 150 mM, Potassium Gluconate 3 mM, Hepes 10 mM, pH 7.42, 300 mOsm) containing 0.5 mg/ml of FLIPR® dye for 30 mins at 37 °C. CFTR function were measured after acute addition of Fsk (10 μM) or 0.01% DMSO control. CFTR functional recordings were measured using the SpectraMax i3. Modulators were then added to stimulate CFTR mediated anion efflux and CFTR mediated fluorescence changes were monitored and imaged every 2 minutes. CFTR channel activity was terminated with addition of CFTR-inh172 (10 μM) and fluorescence changes were monitored every 2 minutes for 5 reads.

Submerge culture apical chloride conductance assay

The ACC assay was used to assess CFTR mediated changes in membrane depolarization using methods as previously described[37]. iPSC-derived airway cultures were replated on collagen IV (Sigma Aldrich, cat# C7521)-coated 96 well plates (Corning, cat# 3603) at AFVE stage Day 3. Cells were dissociated using TrypLE for 3 minutes at 37 °C, and replated at 25,000 cells/well in AFVE media. The AFVE cells continued differentiation until S3B stage. For ACC measurement, the cells on 96 well plates were incubated with zero sodium, chloride and bicarbonate buffer (NMDG 150 mM, Gluconic acid lactone 150 mM, Potassium Gluconate 3 mM, Hepes 10 mM, pH 7.42, 300 mOsm) containing 0.5 mg/ml of FLIPR® dye for 30 mins at 37 °C. CFTR function were measured after acute addition of Fsk (10 μM) or 0.01% DMSO control. CFTR functional recordings were measured using the FLIPR® Tetra High Throughput Cellular Screening System (Molecular Devices), which allowed for simultaneous image acquisition of the entire 96 well plate.

Images were first collected to establish baseline readings over 5 mins at 30 second intervals. Modulators were then added to stimulate CFTR mediated anion efflux and CFTR mediated fluorescence changes were monitored and images were collected at 15 second intervals for 70 frames. CFTR channel activity was terminated with addition of CFTR-Inh172 (10 μM) and fluorescence changes were monitored at 30 second intervals for 25 frames.

Split-open apical chloride conductance (ACC) Assay for CFTR function

The ACC assay was used to assess CFTR mediated changes in membrane depolarization using methods as previously described[53]. Briefly, AO were "split opened" by plating them on 96 well assay plate (Corning; cat# 3603) and allowed to adhere and "hatch" for up to 3 days. At this point, a layer of epithelia can be observed adhered to the transwell membrane. For ACC measurements, the cells on transwells were incubated with zero sodium, chloride and bicarbonate buffer (NMDG 150 mM, Gluconic acid lactone 150 mM, Potassium Gluconate 3 mM, Hepes 10 mM, pH 7.42, 300 mOsm) containing 0.5 mg/ml of FLIPR® dye for 30 mins at 37 °C. CFTR function were measured after acute addition of Fsk (10 μM) or 0.01% DMSO control. CFTR functional recordings were measured using the FLIPR® Tetra High Throughput Cellular Screening System (Molecular Devices), which allowed for simultaneous image acquisition of the entire 96 well plate. Images were first collected to establish baseline readings over 5 mins at 30 second intervals. Modulators were then added to stimulate CFTR mediated anion efflux and CFTR mediated fluorescence changes were monitored and images were collected at 15 second intervals for 70

frames. CFTR channel activity was terminated with addition of CFTR-inh172 (10 μM) and fluorescence changes were monitored at 30 second intervals for 25 frames.

Immunoblotting

Organoids were lysed in modified radioimmunoprecipitation assay (RIPA) buffer (50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH 7.4, 0.2% SDS (Invitrogen, cat# 15553), and 0.1% Triton X-100 (Sigma Aldrich, cat#93443) containing a protease inhibitor cocktail (GBiosciences, cat# 031M-D). Soluble fractions were run by SDS-PAGE on 6% tris-glycine gels (Life Technologies, cat# XP00065BOX). Electrophoresis were run in Tris-glycine-SDS buffer (Wisent, cat# 811570) at 100V for half an hour. Proteins were transferred to nitrocellulose membranes (Bio-Rad) in Tris-glycine buffer (Wisent, cat# 811560) with 20% methanol (Fisher Scientific, cat# 67-56-1) at 100V for one hour, then blocked with 5% milk solution for half an hour at room temperature. Transferred blots were stained with primary antibodies (**Supplementary Table 2**) overnight at 4 °C, followed with secondary antibodies for 1 hour at room temperature. Signal was developed using Clarity Max Western ECL Substrate (Bio-Rad, cat# 1705062) and captured by Licor Odyssey Fc Imaging System. CFTR protein level is normalized to calnexin protein level analyzed by ImageStudio Lite (Li-COR Biosciences).

ACKNOWLEDGEMENT

We would like to thank Dr. D. Kotton for providing the human iPSC reporter lines and Dr. Nagy for the human ESC line. Monoclonal CFTR antibodies #562 were courtesy of J.R. Riordan (University of North Carolina, Chapel Hill, North Carolina, USA). Small molecule CFTRinhibitor-172 was obtained from CFFT. We would like to thank Dr. Christine Bear Lab

(SickKids) for support in CFTR functional and immunoblotting work. This work was supported by the CF Canada-SickKids Program in Individual Therapy (CFIT) (in part) by SickKids CFIT Microsynergy grant (awarded to A.P.W in 2017) and Vertex Pharmaceutical CF Research Innovation Award (awarded to A.P.W in 2017), the Genome Canada and the Ontario Genomics Institute (OGI-148), the SickKids Foundation & CIHR-IHDCHY grant (NI20-1070), and the Stem Cell Network Early Career Investigator - Innovation Award (FY21/ECI23 Wong). Computational power was enabled by support provided by Compute Canada (www.computecanada.ca) for processing our scRNA-seq datasets. We would like to thank Drs. Chris McGinnis and Zev J Gartner (University of California San Francisco, U.S.A) Lab for providing the reagents and protocols for MULTI-seq experiment.

AUTHOR CONTRIBUTIONS

S.Y.N, H.Q, J.D, J.L, E.H, A.P.W performed experimental assays, analyzed the data, and contributed to the preparation of the manuscript. S.Y.N, and H.Q contributed equally to the work. A.P.W designed the study, interpreted the data and wrote the manuscript. All authors read and approved the manuscript.

DISCUSSION

Human PSCs have been an invaluable resource in generating human lung tissues for regenerative medicine and development. While there are many protocols available to date that seemingly uses common soluble factors to induce directed differentiation of hPSC to lung epithelial cells, incremental improvements have led to the ability of generating distal versus proximal lung epithelial cell types, mature cell types, and renewable progenitors[13–15]. The latter is especially important for generating a renewable source of lung epithelial cells from iPSC without having to repeat a lengthy and costly differentiation protocol from PSC state. We previously were the first to report a method to generate airway epithelial cells from human pluripotent stem cells that can be used to model Cystic fibrosis disease *in-vitro*[18]. However, without fractionation or enrichment for putative lung progenitor cells during the early stages of differentiation, non-lung epithelial lineages co-emerged throughout the differentiation process, not surprisingly since many of the morphogenetic pathways are shared during development. Few groups have shown efficient generation of NKX2-1+ lung progenitor (>90%)[26,27], while others have relied on using reporter strategies (NKX2-1^{GFP+}, NKX2-1^{GFP+}SFTPC$^{tdTomato+}$, NKX2-1^{GFP+}P63$^{tdTomato+}$ [13–15]or identified putative cell surface markers (CD47^{high} and CD26^{low} or CPM^{+})[14,24,25,36] to isolate or enrich for the lung progenitor population. Nonetheless, significant improvements in purifying lung epithelial progenitor cells have provided an enriched source of cells to model lung diseases *in-vitro*.[11,13–15] Temporal activation of Wnt pathway is key in driving distal type II alveolar epithelial cell (AEC2)[14,15] and can be used to model distal respiratory diseases such as surfactant deficiencies. Inhibition of the pathway promotes proximal cell differentiation and can be used to generate proximal epithelial organoids to model CF airway disease *in-vitro*.[14]

Recently, Hawkins et al[13] showed a robust method of generating renewable sources of basal stem cells with the capacity to differentiate into multi-epithelial cell types including secretory and ciliated cells *in-vitro* and repopulate the denuded rat trachea *in-vivo*. Using single cell analyses of primary fetal lung tissue of the pseudoglandular stage, Miller et al identified SMAD signalling as a key regulator of distal tip-to-airway transition of the epithelial cells and in doing so can generate renewable fetal basal stem cells with multilineage differentiation potential[11]. Here, we present a similar approach of generating a renewable progenitor population enriched in basal stem cells that improves the differentiation efficiency in generating fetal-like lung cells and mature airway epithelia in ALI. Similar to previous protocols, dual SMAD inhibition using small molecules A8301 and DMH1[13] in addition CHIR99021 (Wnt pathway activation), and all-trans retinoic acid (RA pathway activation)[25,26] generates epithelial organoids with expansion potential. Employing single cell RNA sequencing to benchmark and determine the heterogeneity of the cell types in our differentiation cultures, we identify unique clusters of lung cells found in the developing fetal lung[31], as well as the co-emergence of mesenchymal cells. The origin of the mesenchymal cell population remains unclear however we speculate that epithelial-mesenchymal transition may contribute to these cell populations[41]. Indeed, gene ontology assessment of differentially expressed genes (DEG) of each cell cluster revealed high activity of epithelial to mesenchymal transition (EMT and myofibroblast) and positive regulation of epithelial to mesenchymal transition (myofibroblast and POSTN2^{hi} mesenchymal) in the top GO terms.

Limited access to primary human fetal tissues for biomedical research has made it particularly difficult to study cellular and molecular mechanisms driving human lung development. Much of what we know about lung development has traditionally been extrapolated from heralding efforts in animal models[4,54,55] and only recently been studied in limited human

tissue samples[31,56]. These models have been useful in describing conserved developmental pathways however not all pathways are conserved and human disease development is not always recapitulated in animal models such as CF[57]. Human stem cell-derived models have gained increasing traction to study human-specific developmental pathways and disease modeling[27,36,58–61]. As the differentiation of hPSC follows a directed developmental trajectory, these new human tissue models are of great interest to study fundamental mechanisms driving lung cell lineage formation. However, many of the hPSC-derived "recapitulated" developmental stages have largely been validated with few gene expression markers benchmarked from previous mouse studies including our previous developmental protocol[18,22]. Few studies have employed human primary tissue to validate stage-specific match of the cells formed from *in-vitro* differentiation using RNA sequencing (single cell or bulk) of *select* primary tissues[11,15]. Current challenges of *in-vitro* human pluripotent stem cell (hPSC)-derived differentiations towards tissue-specific cell types is the faithful modeling of *in-vivo* development. In our study, we generate FL cells reminiscent of pseudoglandular fetal lung that contain cells with unique SOX2 and SOX9 expression patterns similar to human fetal bud tip proximal-distal patterning[31]. This provides a unique hPSC-derived lung model to assess cell patterning and specification during fetal lung development and potentially be used to determine genetic causes of congenital pulmonary airway malformations[62]. This is particularly important as access to human fetal tissues for research is limited and confounded by the scarcity of fetal lungs with congenital abnormalities. As a case in point, we confirm the expression of CFTR, a chloride channel protein implicated in CF disease, in our FL cultures during this stage that has previously been found to be up-regulated during pseudoglandular lung development[32]. We demonstrated measurable CFTR function in FL cultures which suggest a developmental role of CFTR in fetal lung cells that may not be appreciated. Thus, we have now

identified a potential opportunity to understand the congenital cause of CF lung diseases using hPSC-derived FL models where previous studies have found pathological manifestations of CF diseases in newborns[63–68].

A new advantage of our improved differentiation protocol is the generation of expandable hPSC-derived AO from FL cells. Similar to previous findings[11,13], dual SMAD inhibition appears to be enriched for basal stem cell-like cells expressing TP63 capable of multilineage differentiation *in-vitro*. Importantly, hPSC-derived AO provides a user-friendly opportunity that is time and cost-effective for researchers with no experience or capacity to work with hPSC. Their proliferative and relatively undifferentiated nature enables rapid expansion of AO for regenerative studies[13] and developing high throughput assays for disease models and/or drug screens[37]. Upon differentiation, the airway epithelial organoids recapitulate proximal airway phenotype with basal cells on the apical side with mature cell types (secretory, ciliatory) and CFTR localized to the luminal side. While some studies have suggested airway organoids can be used in forskolin-induced swelling (FIS) assays to assess CFTR function[14], the slow (>4 hours of FIS) and often minimal swelling (often <10%) observed in control AO, likely due to much lower expression of CFTR limits the use of AO in rapid screens FIS screens of CFTR modulators, unlike intestinal organoids where CFTR is more abundantly expressed[30,69]. Instead, our open AO apical chloride conductance (ACC) assay is much more sensitive at detecting CFTR-mediated responses and offers an alternative strategy to measure CFTR modulator responses in CF iPSC-derived AO. It should be noted that an additional limitation of AO is the relatively simple epithelium that is formed devoid of other airway epithelial cells types including rare cell types neuroendocrine cells and ionocytes. Therefore, contributions of the rare ionocytes in CF disease and responses to therapies are not captured with the AO model.

Similar to Hawkins et al[13], further differentiating the cells in ALI as we previously described[18,22], induces maturation and polarization of the epithelium such that by the end of 5 weeks in ALI, a pseudostratified epithelium is formed, reminiscent of the large airways. Our hPSC-derived ALI culture showed an improved level of CFTR protein and measurable function (up to 40% function above baseline). Most importantly, air-liquid interface mimics post-natal airway development. In our study, we used scRNAseq to reveal transitory cell types that could not be captured previously[22]. Many of the known airway cell types including ciliated, basal, goblet, club, PNEC and ionocytes were identified in our ALI cultures, in addition to a large population of mesenchymal cells co-developed during cellular differentiation. Interestingly, an unknown cluster of cells appear to emerge from cells undergoing EMT and express some genes (*IQGAP2, ID3, ID4*) previously associated with ionocytes but not *FOXI1*[8]. Moreover, CFTR is not highly expressed in this cluster of cells suggesting these may be pre-ionocyte cell types and that our current differentiation protocol does not support ionocyte cell development.

Previously, pulmonary cell types were identified through a snapshot measuring the most abundant genes at each stage of the differentiation protocol[18,22]. In our current integrated analyses of the FL and ALI differentiation stages through functional pseudotime analysis, a trajectory inference method that models gene expression changes as a function of pseudotime along lineages[52,70], we were able to identify lung cell lineage hierarchies and relationships as they differentiate in culture. We found previously identified lineage relationships such as basal cells as the source of secretory and ciliated cells[71]. We also identified the emergence of the embryonic bud (NKX2-1^{hi}SOX9^{hi}SOX2^{lo}) that gives rise to stalk (NKX2-1^{hi}SOX9^{lo}SOX2^{hi}) progenitor cell types and the presence (NKX2-1^{hi}SOX9^{lo}SOX2^{lo}) as previously found[31,71]. Interestingly, we also identified ciliated cells as a source of mesenchymal cells. Downregulation of CFTR has previously

been associated with down-regulation of e-cadherin which induced EMT phenotypic changes in cancer and CF[2,72]. Interestingly, we also found overlaps of cell clusters such as NKX2-1^{hi}SOX9^{hi}SOX2^{lo} lung progenitors with ALI ciliated cells and ALI cycling basal cells with FL ciliated precursors, suggesting common transcriptional programs shared in mature and FL cells. An isolated cluster of PNEC that emerged in the differentiation shared no common gene expression to other FL cells and may have emerged from a separate ancestral lineage that has not been captured with our dataset[73]. Overall, we are the first to show unique lineage relationships between the cells as they emerge in our differentiation of PSC to mature airway cells from the fetal to mature cell states.

Overall, we present a new and improved hPSC differentiation protocol that models human lung cell development *in-vitro* and may be used to study early developmental mechanisms driving congenital lung diseases. Furthermore, creating new pre-clinical tools to investigate signalling mechanisms driving development, that can reactivate during airway regeneration, can provide insight into activating novel endogenous pathways for cell replacement therapy.

Figure Legends

Fig 1. Refined *in-vitro* step-wise differentiation of human pluripotent stem cells to generate fetal lung (Stage 3) cells, immature organoids (Stage 4) and mature ALI cultures (Stage 5). *Top*: Renewable fetal lung organoids can be generated from hPSC-derived fetal lung (Stage 3) cells and further differentiated to generate immature lung organoids (Stage 4) that can be dissociated onto transwells to culture in air liquid interface (ALI) for at least 5 weeks. *Bottom:* Detailed timeline comparing the refined differentiation method to our previously published method[22].

Fig 2. Stage-matching *in-vitro* differentiated hPSC-derived fetal lung (FL) cells to *in-vivo* human fetal lung tissue. ***A.*** Differentiation to stage 3 fetal lung (FL) cells. ***B.*** Pearson correlation analysis of upregulated genes between hPSC-derived fetal lung cells and primary fetal samples from pseudoglandular (GW 16) and early-canalicular (GW18) shows hPSC-derived FL cells are highly correlated with pseudoglandular fetal lung tissues (GW16) and early canalicular fetal lung tissues (GW18) ($p > 0.85$). Reference map shows the Pearson correlation coefficient range from 0.855 to 0.924. ***C.*** Single cell RNA sequencing of *in-vitro* differentiated FL cells for both new and old protocols. A total of 3000 cells (per protocol) were analyzed, and the single cell transcriptome was visualized by the UMAP plots of unsupervised clustering of cells for both new and old protocols. ***D.*** Average expression level for each gene and frequency of cells expressing the gene is shown by dotplot of FL cells derived from the new protocol. Reference map shows intensity of the average gene expression level and percentage of specific gene-expressing cells are demarcated by size of the circles. ***E.*** Feature plots of select gene expression show distribution of the expression in the UMAP clusters. Reference bar represents the avg_log2FC. Minimum value set as 0.2. ***F.***

The frequency of CFTR-expressing cells in all clusters. ***G.*** Heatmap represents the top (up to) 5 and cluster-specific gene ontology (GO) terms identified with each cluster based on differential gene expression (DEG). The Benjamini-adjusted p-value in log scale (<0.05 or log (-1.3)) was used to determine DEG. Empty boxes represent no or insignificant p-values. Note: each cluster is colour coded based on the UMAP clusters in **C**.

Fig 3. Phenotype and functional validation of *in-vitro* differentiated fetal lung (FL) cells . ***A.*** Immunofluorescence staining of FL cells for proteins highly expressed in the developing fetal lung tissue TTF1, SOX2, SOX9, CFTR and beta-catenin. Two human pluripotent stem cell lines (CA1 and BU3-NKX2-1^{GFP}) were used. Non-immune isotype controls were used to determine non-specific binding of secondary antibodies. White bar represents 41 microns (low magnification) and 20 microns (high magnification) respectively. ***B.*** Flow cytometric analysis demonstrating improved generation of lung (TTF1+) cells expressing the basal stem cell marker (ΔNP63) in the new and old protocol. **C,D.** Representative apical chloride conductance traces of FL cells demonstrate robust CFTR functional response to forskolin-induced channel activity (red) that is inhibited by CFTR inhibitor-172 (n=5) in old and new protocols, respectively. ***E.*** Quantitative box-plots comparing maximal forskolin-induced CFTR activation in independent sets of FL cells from the old and new differentiation protocols.

Fig 4. Phenotype and expansion characterization of *in-vitro* proliferative lung epithelial organoids. ***A.*** Differentiation to FL airway epithelial progenitor organoids. ***B.*** Percentage of CD326^{+} expressing cells in fetal lung cells assessed by magnetic-activated cell sorting assay. ***C.*** Representative DIC image of airway epithelial organoids at passage 1 and 4. ***D.*** Expansion curve

of organoids of number organoids per dome. ***E.*** Number of epithelial organoids generated from 3 different hPSC lines and form from fetal lung organoids in differentiation medium or directly from Stage 4 cells (Fig 1). ***F.*** Efficiency of generating epithelial organoids range from 30-85% depending on the line. ***G.*** Representative images of BU3-NKX2-1^{GFP}/TP63^{dTomato} in DIC and immunofluorescence imaging for TP63^{tdTomato} reporter. ***H.*** Flow cytometry analysis showed high levels of basal (ΔNP63) and epithelial (CD326) in stage 3 organoids. Red and blue contour plots represent immunostained and unstained cells, respectively. ***I.*** Immunofluorescence images shows undifferentiated lung epithelial organoids expresses fetal lung (TTF1, SOX2, SOX9) epithelial (KRT8/18) and pulmonary cell types (ΔNP63, Basal stem cell; KRT14, Basal cells; SCGB3A2, secretory cells). Rabbit primary antibodies shown in green, mouse primary antibodies in red and goat or rat primary antibodies shown in white. Non-immune reactive isotypes (IgG) were used as control staining. White scale bar is automatically generated and represents 41 microns (low magnification) and 20 microns (high magnification) images.

Fig 5. Phenotype and functional characterization of *in-vitro* differentiated lung epithelial organoids. ***A.*** Differentiation of airway epithelial organoids. ***B.*** Representative DIC images of *in-vitro*-derived airway epithelial organoids and primary bronchial cells-derived airway organoids. ***C.*** Characterization of *in-vitro*-derived airway epithelial organoids for lung airway (SOX2, TTF1) pan-epithelial (panCK, E-cadherin) and specific epithelial cell markers (ZO1, tight junction; acetylated alpha tubulin, cilia on ciliated cells; MUC5AC, Goblet cell; ΔNP63, Basal stem cell; KRT14, Basal cell; SCGB1A1, Club cell). Epithelial organoids express mature luminal epithelial markers KRT8/18 and luminal expression of chloride channel CFTR. Notably, pan-endoderm transcription factor FOXA2 is weakly expressed in a subset of cells in the organoids. ***D.*** Heatmap

of gene expression analysis of epithelial organoids in undifferentiated (stage 3 organoids) and differentiated (stage 4 organoids) show an increase in expression of epithelial-specific lineage markers (*TRP63, K5, FOXJ1, K8, MUC16*) and a near 2-fold increase in *CFTR* expression. Genes were normalized to housekeeping gene *GAPDH* and expressed relative to RNA from trachea (expression level = 1.0). At least 3 technical replicates were performed, n= 3 biological cell lines. ***E.*** Only epithelial ($CD326^+$) cells (sorted by magnetic activated cell sorting (MACS) formed organoids expressing the mature CFTR protein (Band C). Anti-calnexin was used as a control for protein loading. ***F.*** Representative ACC trace of CFTR function upon stimulation with forskolin and inhibited with CFTR inhibitor-172. ***G.*** Box-plot demonstrates the average maximum forskolin-induced response level of differentiated organoids from different sets of organoids. White scale bar is automatically generated and represents 41 microns.

Fig 6. Single cell profiling of *in-vitro*-derived mature ALI airway cells compared to human bronchial epithelial cultures. *A*. Differentiation to mature airway epithelial cells in ALI for 5 weeks demonstrate. ***B***. A total of *in-vitro* ~3000 cells per protocol were analyzed and the single-cell transcriptome was visualized by UMAP plot. Unsupervised clustering reveals 11 and 12 cell clusters in the old and new differentiation protocols, respectively. ***C.*** Dot plot shows the average gene expression of pulmonary cell-specific genes in each cluster for ALI cultures derived from the new protocol. Reference map shows the range of gene expression level and the number of cells in each cluster is represented by circle size. ***D.*** UMAP presents the average gene expression of selected genes in each cell. Reference bar represents the avg_log2FC. Minimum value set as 0.2. ***E.*** Frequency of CFTR-expressing cells in each cluster of mature ALI airway cells. ***F.*** Heatmap represents the top (up to) 5 gene ontology (GO) terms identified with each cluster based on

differential gene expression (DEG). The Benjamini-adjusted p-value in log scale (<0.05 or log (-1.3)) was used to determine DEG. Empty boxes represent no or insignificant p-values. Note: each cluster is colour coded based on the UMAP clusters in **B**.

Fig 7. Phenotype and functional validation of *in-vitro* differentiated mature airway epithelial cells. ***A*** and ***A'.*** Immunofluorescence staining of ALI cultures (low and high magnification, respectively) from Stage 5 cells show proximal (SOX2) airway (TTF1) phenotype which is reminiscent of a pseudostratified structure in control lung tissue ***A''***. ***B*** and ***B'.*** Basal (P63), secretory (SCGB3A2) and ciliated (acetylated alpha-tubulin) cells are present and form a polarized epithelial similar to control lung airway ***B''. C.*** Non-immune isotype controls were used to determine non-specific binding of secondary antibodies. White bar represents 41 microns and 20 microns for low and high magnification, respectively. ***D.*** Immunoblot analysis for CFTR protein species shows higher levels of a mature CFTR protein (Band C, ~160kDa) in stage 5 cells in the new differentiation to mature airway epithelia derived from both human PSC lines. Anti-calnexin was stained to evaluate protein loading. Human bronchial epithelial cells (HBE WT-CFTR) expressing wild-type CFTR were used as positive control whereas CFTR knock-out in human bronchial epithelial cells (HBE KO-CFTR) was used as negative control. ***E.*** Representative apical chloride conductance trace of mature lung cells demonstrating robust CFTR functional response to forskolin-induced channel activity (red) that is inhibited by CFTR inhibitor-172 (n=5). ***F.*** Comparing maximal forskolin-induced CFTR activation of stage 5 cells demonstrates a 2-fold higher level of functional activity in the new differentiation protocol.

Fig 8. Pseudotime cell trajectory of differentiated cells from fetal to mature states.

A. Integrated UMAP of fetal lung (FL) and mature (ALI) cells calculated based on common anchor genes. ***B.*** UMAP visualization of partitioned clusters to determine lineage trajectories of cells. ***C.*** UMAP visualization of the pseudotime trajectories in each partition. **D.** Predicted lineage trajectories based on differentiation progression.

References

1. Herriges, M. & Morrisey, E. E. Lung development: orchestrating the generation and regeneration of a complex organ. *Development* **141**, 502–513 (2014).

2. Amaral, M. D., Quaresma, M. C. & Pankonien, I. What Role Does CFTR Play in Development, Differentiation, Regeneration and Cancer? *Int. J. Mol. Sci.* **21**, 3133 (2020).

3. Morrisey, E. E. & Hogan, B. L. M. Preparing for the First Breath: Genetic and Cellular Mechanisms in Lung Development. *Dev. Cell* **18**, 8–23 (2010).

4. Treutlein, B. *et al.* Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. (2014) doi:10.1038/nature13173.

5. Travaglini, K. J. *et al.* A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619–625 (2020).

6. Du, Y. *et al.* Lung Gene Expression Analysis (LGEA): an integrative web portal for comprehensive gene expression data analysis in lung development. *Thorax* **72**, 481–484 (2017).

7. Montoro, D. T. *et al.* A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319–324 (2018).

8. Plasschaert, L. W. *et al.* A single-cell atlas of the airway epithelium reveals the the CFTR-rich pulmonary ionocyte. *Nature* (2018) doi:10.1038/s41586-018-0394-6.

9. He, M. *et al.* Chloride channels regulate differentiation and barrier functions of the mammalian airway. *Elife* **9**, (2020).

10. Carraro, G. *et al.* Transcriptional analysis of cystic fibrosis airways at single-cell resolution reveals altered epithelial cell states and composition. *Nat. Med.* **27**, 806–814 (2021).

11. Miller, A. J. *et al.* In Vitro and In Vivo Development of the Human Airway at Single-Cell Resolution. *Dev. Cell* **53**, 117-128.e6 (2020).

12. Rodrigues Toste de Carvalho, A. L. *et al.* The in vitro multilineage differentiation and maturation of lung and airway cells from human pluripotent stem cell–derived lung progenitors in 3D. *Nat. Protoc.* **16**, 1802–1829 (2021).

13. Hawkins, F. J. *et al.* Derivation of Airway Basal Stem Cells from Human Pluripotent Stem Cells. *Cell Stem Cell* **28**, 79-95.e8 (2021).

14. McCauley, K. B. *et al.* Efficient Derivation of Functional Human Airway Epithelium from Pluripotent Stem Cells via Temporal Regulation of Wnt Signaling. *Cell Stem Cell* **20**, 844-857.e6 (2017).

15. Jacob, A. *et al.* Differentiation of Human Pluripotent Stem Cells into Functional Lung Alveolar Epithelial Cells. *Cell Stem Cell* **21**, 472-488.e10 (2017).

16. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663–676 (2006).

17. Takahashi, K. *et al.* Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* **131**, 861–872 (2007).

18. Wong, A. P. *et al.* Directed differentiation of human pluripotent stem cells into mature airway epithelia expressing functional CFTR protein. *Nat. Biotechnol.* **30**, 876–882

(2012).

19. d'Souza, S. L. *et al.* Generation of anterior foregut endoderm from human embryonic and induced pluripotent stem cells. *Nat. Biotechnol.* (2011) doi:10.1038/nbt.1788.

20. Mou, H. *et al.* Generation of multipotent lung and airway progenitors from mouse ESCs and patient-specific cystic fibrosis iPSCs. *Cell Stem Cell* **10**, 385–397 (2012).

21. Firth, A. L. *et al.* Generation of multiciliated cells in functional airway epithelia from human induced pluripotent stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **111**, (2014).

22. Wong, A. P. *et al.* Efficient generation of functional CFTR-expressing airway epithelial cells from human pluripotent stem cells. *Nat. Protoc.* **10**, 363–381 (2015).

23. Ahmadi, S. *et al.* Phenotypic profiling of CFTR modulators in patient-derived respiratory epithelia. *npj Genomic Med.* **2**, 12 (2017).

24. Konishi, S. *et al.* Directed Induction of Functional Multi-ciliated Cells in Proximal Airway Epithelial Spheroids from Human Pluripotent Stem Cells. *Stem Cell Reports* **6**, 18–25 (2016).

25. Gotoh, S. *et al.* Generation of alveolar epithelial spheroids via isolated progenitor cells from human pluripotent stem cells. *Stem cell reports* **3**, 394–403 (2014).

26. Huang, S. X. L. *et al.* Efficient generation of lung and airway epithelial cells from human pluripotent stem cells. *Nat. Biotechnol.* **32**, 84–91 (2014).

27. Chen, Y.-W. *et al.* A three-dimensional model of human lung development and disease from pluripotent stem cells. *Nat. Cell Biol.* **19**, 542–549 (2017).

28. McCauley, K. B., Hawkins, F. & Kotton, D. N. Derivation of Epithelial-Only Airway Organoids from Human Pluripotent Stem Cells. *Curr. Protoc. Stem Cell Biol.* **45**, (2018).

29. Eckford, P. D. W. *et al.* The CF Canada-Sick Kids Program in individual CF therapy: A resource for the advancement of personalized medicine in CF. *J. Cyst. Fibros.* **18**, 35–43 (2019).

30. Berkers, G. *et al.* Rectal Organoids Enable Personalized Treatment of Cystic Fibrosis. *Cell Rep.* **26**, 1701-1708.e3 (2019).

31. Nikolic, M. Z. *et al.* Human embryonic lung epithelial tips are multipotent progenitors that can be expanded in vitro as long-term self-renewing organoids. (2017) doi:10.7554/eLife.26575.001.

32. Marcorelles, P., Montier, T., Gillet, D., Lagarde, N. & Ferec, C. Evolution of CFTR protein distribution in lung tissue from normal and CF human fetuses. *Pediatr. Pulmonol.* **42**, 1032–1040 (2007).

33. Ruiz García, S. *et al.* Novel dynamics of human mucociliary differentiation revealed by single-cell RNA sequencing of nasal epithelial cultures. *Development* **146**, (2019).

34. Sunday, M. E. Pulmonary neuroendocrine cells and lung development. *Endocr. Pathol.* **7**, 173–201 (1996).

35. Aros, C. J., Pantoja, C. J. & Gomperts, B. N. Wnt signaling in lung development, regeneration, and disease progression. *Commun. Biol.* **4**, 601 (2021).

36. Hawkins, F. *et al.* Prospective isolation of NKX2-1-expressing human lung progenitors derived from pluripotent stem cells. *J. Clin. Invest.* **127**, 2277–2294 (2017).

37. Jiang, J. X. *et al.* A new platform for high-throughput therapy testing on iPSC-derived , immature airway from Cystic Fibrosis Patients. *bioRxiv* 1–22 (2021) doi:10.1101/2021.07.05.451013.

38. Tizzano, E. F., O'Brodovich, H., Chitayat, D., Bènichou, J. C. & Buchwald, M. Regional expression of CFTR in developing human respiratory tissues. *Am. J. Respir. Cell Mol. Biol.* **10**, 355–362 (1994).

39. Schittny, J. C. Development of the lung. *Cell Tissue Res.* **367**, 427–444 (2017).

40. Tsukui, T. *et al.* Collagen-producing lung cell atlas identifies multiple subsets with distinct localization and relevance to fibrosis. *Nat. Commun.* **11**, 1920 (2020).

41. Zhou, Y. *et al.* Extracellular matrix in lung development, homeostasis and disease. *Matrix Biol.* **73**, 77–104 (2018).

42. Curtis, J. L. Cell-mediated adaptive immune defense of the lungs. *Proc. Am. Thorac. Soc.* **2**, 412–6 (2005).

43. Kaiser, F., Huebecker, M. & Wachten, D. Sphingolipids controlling ciliary and microvillar function. *FEBS Lett.* **594**, 3652–3667 (2020).

44. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

45. Wolf, F. A. *et al.* PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59 (2019).

46. Mcginnis, C. S. *et al.* MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. doi:10.1038/s41592-019-0433-8.

47. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species A n A ly s I s. *Nat. Biotechnol. Vol.* **36**, (2018).

48. Soneson, C. & Robinson, M. D. Bias, robustness and scalability in single-cell differential expression analysis. *Nat. Methods* **15**, 255–261 (2018).

49. Hurley, K. *et al.* Reconstructed Single-Cell Fate Trajectories Define Lineage Plasticity Windows during Differentiation of Human PSC-Derived Distal Lung Progenitors. *Cell Stem Cell* **26**, 593-608.e8 (2020).

50. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).

51. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).

52. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).

53. Xia, S. *et al.* High-throughput functional analysis of CFTR and other apically localized channels in iPSC derived intestinal organoids. *bioRxiv* (2021).

54. Metzger, R. J., Klein, O. D., Martin, G. R. & Krasnow, M. A. The branching programme of mouse lung development. *Nature* **453**, 745–750 (2008).

55. Lavelle, G. M., White, M. M., Browne, N., McElvaney, N. G. & Reeves, E. P. Animal Models of Cystic Fibrosis Pathology: Phenotypic Parallels and Divergences. *Biomed Res.*

Int. **2016**, 5258727 (2016).

56. Miller, A. J. *et al.* In Vitro Induction and In Vivo Engraftment of Lung Bud Tip Progenitor Cells Derived from Human Pluripotent Stem Cells. *Stem Cell Reports* **10**, 101–119 (2018).

57. Qian, X. *et al.* Brain-Region-Specific Organoids Using Mini-bioreactors for Modeling ZIKV Exposure. *Cell* **165**, 1238–1254 (2016).

58. Garcez, P. P. *et al.* Zika virus impairs growth in human neurospheres and brain organoids. *Science (80-.).* **352**, 816–818 (2016).

59. Sampaziotis, F. *et al.* Cholangiocytes derived from human induced pluripotent stem cells for disease modeling and drug validation. *Nat. Biotechnol.* **33**, 845–852 (2015).

60. Itzhaki, I. *et al.* Modelling the long QT syndrome with induced pluripotent stem cells. *Nature* **471**, 225–230 (2011).

61. Lancaster, M. A. *et al.* Cerebral organoids model human brain development and microcephaly. doi:10.1038/nature12517.

62. Taylor, B., Rice, A., Nicholson, A. G., Hind, M. & Dean, C. H. Mechanism of lung development in the aetiology of adult congenital pulmonary airway malformations. *Thorax* **75**, 1001–1003 (2020).

63. Khan, T. Z. *et al.* Early pulmonary inflammation in infants with cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **151**, 1075–82 (1995).

64. Ornoy, A. *et al.* Pathological confirmation of cystic fibrosis in the fetus following prenatal diagnosis. *Am. J. Med. Genet.* **28**, 935–47 (1987).

65. Tepper, R. S. *et al.* Infants with cystic fibrosis: pulmonary function at diagnosis. *Pediatr. Pulmonol.* **5**, 15–8 (1988).

66. Ackerman, V., Montogomery, G., Eigen, H. & Tepper, R. S. Assessment of airway responsiveness in infants with cystic fibrosis. *Am. Rev. Respir. Dis.* **144**, 344–6 (1991).

67. Beardsmore, C. S. *et al.* Lung function in infants with cystic fibrosis. *Thorax* **43**, 545–51 (1988).

68. Verhaeghe, C., Tabruyn, S. P., Oury, C., Bours, V. & Griffioen, A. W. Intrinsic pro-angiogenic status of cystic fibrosis airway epithelial cells. *Biochem. Biophys. Res. Commun.* **356**, 745–9 (2007).

69. Fleischer, A. *et al.* iPSC-Derived Intestinal Organoids from Cystic Fibrosis Patients Acquire CFTR Activity upon TALEN-Mediated Repair of the p.F508del Mutation. *Mol. Ther. - Methods Clin. Dev.* **17**, 858–870 (2020).

70. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

71. Nikolić, M. Z., Sun, D. & Rawlins, E. L. Human lung development: recent progress and new challenges. *Development* (2018) doi:10.1242/dev.163485.

72. Li, J. *et al.* The cystic fibrosis transmembrane conductance regulator as a biomarker in non-small cell lung cancer. *Int. J. Oncol.* **46**, 2107–15 (2015).

73. Mou, H. *et al.* Airway basal stem cells generate distinct subpopulations of PNECs. *Cell Rep.* **35**, 109011 (2021).

Figure 1

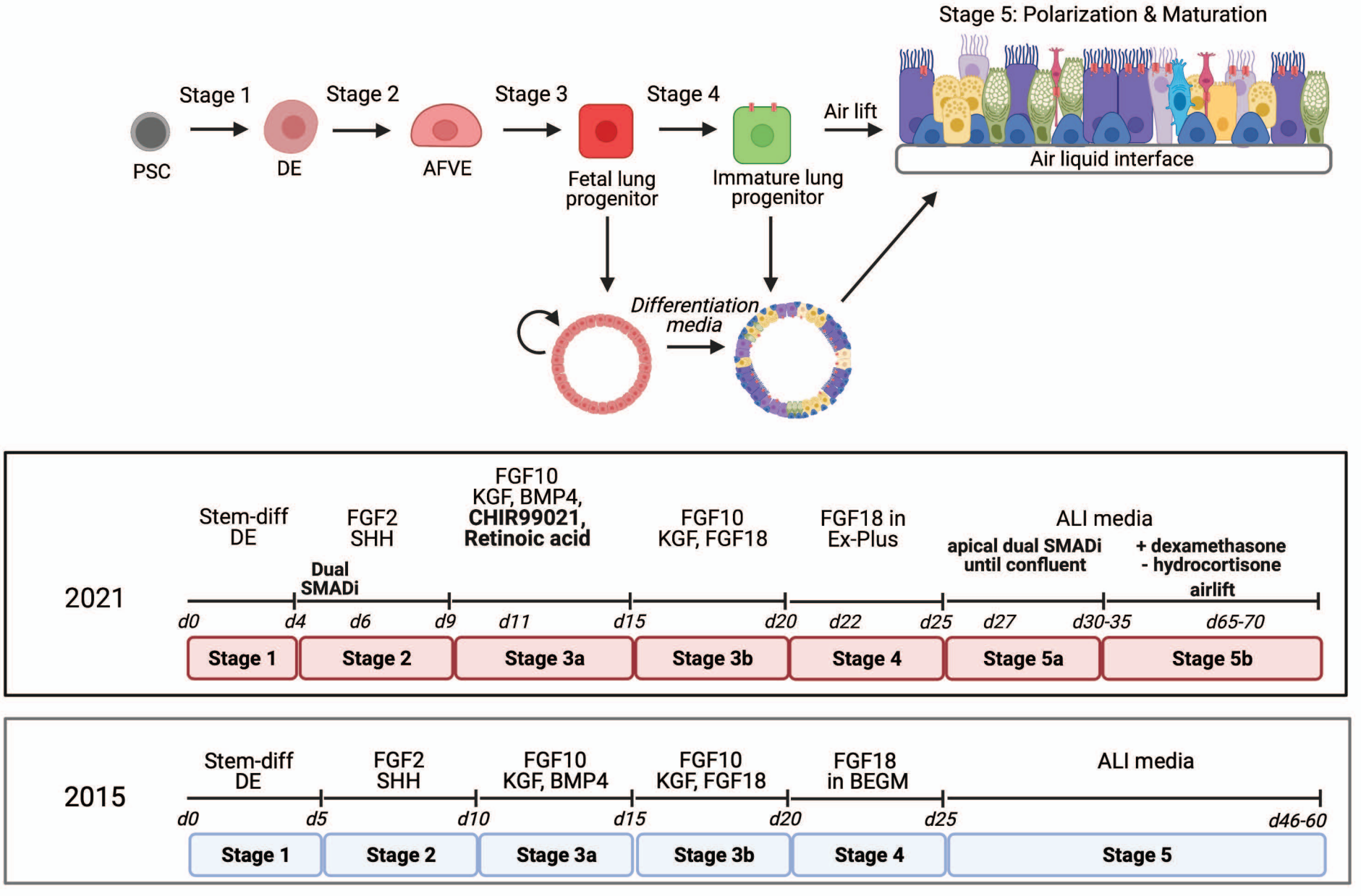

Stage 5: Polarization & Maturation
Stage 1
Stage 2
Stage 3
Stage 4
Air lift
PSC
DE
AFVE
Fetal lung progenitor
Immature lung progenitor
Air liquid interface
Differentiation media
2021
Stem-diff DE
FGF2 SHH
Dual SMADi
FGF10 KGF, BMP4, CHIR99021, Retinoic acid
FGF10 KGF, FGF18
FGF18 in Ex-Plus
ALI media
apical dual SMADi until confluent
+ dexamethasone - hydrocortisone
airlift
d0
d4
d6
d9
d11
d15
d20
d22
d25
d27
d30-35
d65-70
Stage 1
Stage 2
Stage 3a
Stage 3b
Stage 4
Stage 5a
Stage 5b
2015
Stem-diff DE
FGF2 SHH
FGF10 KGF, BMP4
FGF10 KGF, FGF18
FGF18 in BEGM
ALI media
d0
d5
d10
d15
d20
d25
d46-60
Stage 1
Stage 2
Stage 3a
Stage 3b
Stage 4
Stage 5

Figure 2

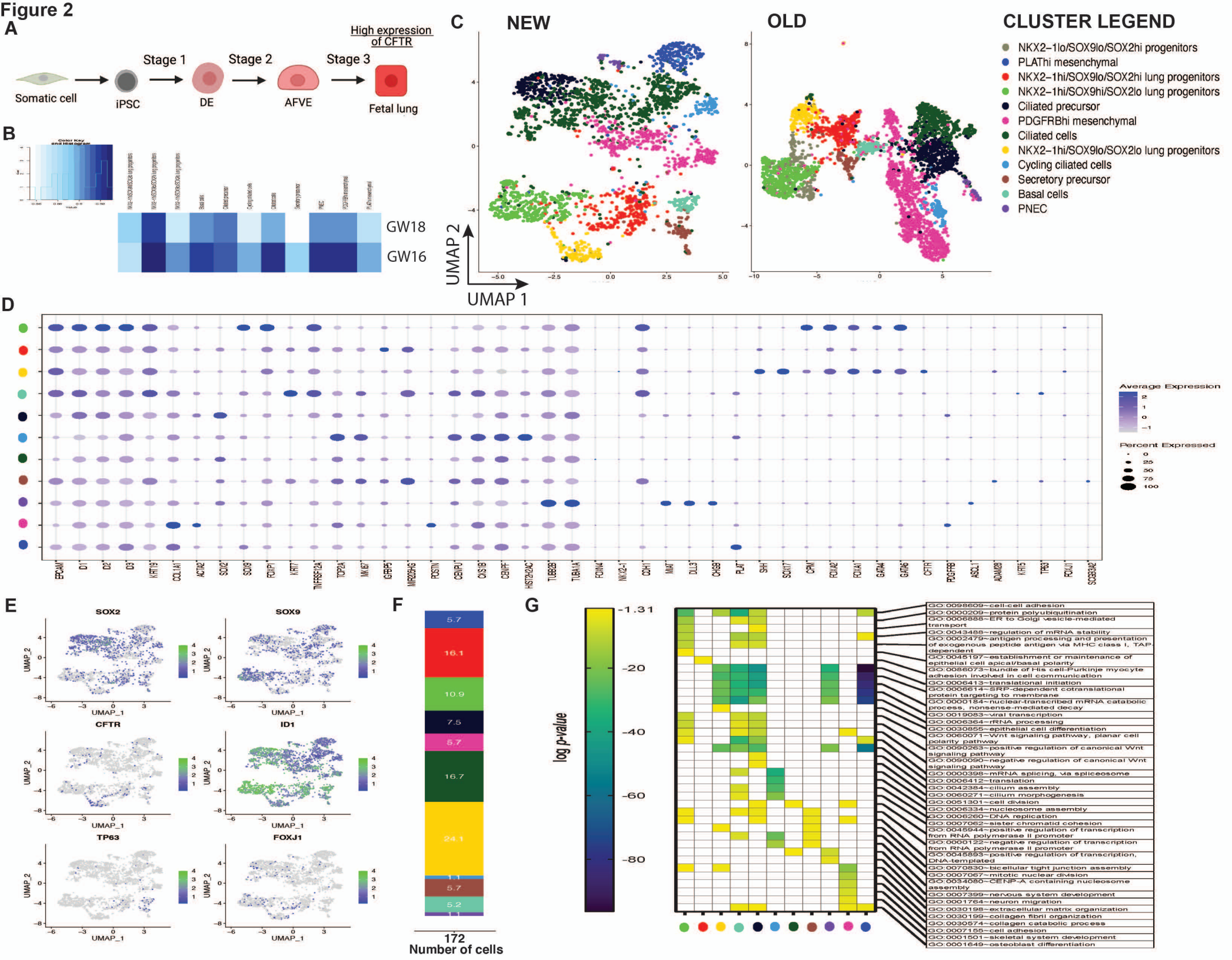

A
High expression of CFTR
Stage 1
Stage 2
Stage 3
Somatic cell
iPSC
DE
AFVE
Fetal lung
B
GW18
GW16
C
NEW
OLD
UMAP 2
UMAP 1
CLUSTER LEGEND
NKX2-1lo/SOX9lo/SOX2hi progenitors
PLAThi mesenchymal
NKX2-1hi/SOX9lo/SOX2hi lung progenitors
NKX2-1hi/SOX9hi/SOX2lo lung progenitors
Ciliated precursor
PDGFRBhi mesenchymal
Ciliated cells
NKX2-1hi/SOX9lo/SOX2lo lung progenitors
Cycling ciliated cells
Secretory precursor
Basal cells
PNEC
D
Average Expression
Percent Expressed
E
SOX2
SOX9
CFTR
ID1
TP63
FOXJ1
UMAP_1
UMAP_2
F
172
Number of cells
G
log p-value
-1.31
-20
-40
-60
-80
GO:0098609~cell-cell adhesion
GO:0000209~protein polyubiquitination
GO:0006888~ER to Golgi vesicle-mediated transport
GO:0043488~regulation of mRNA stability
GO:0002479~antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent
GO:0045197~establishment or maintenance of epithelial cell apical/basal polarity
GO:0086073~bundle of His cell-Purkinje myocyte adhesion involved in cell communication
GO:0006413~translational initiation
GO:0006614~SRP-dependent cotranslational protein targeting to membrane
GO:0000184~nuclear-transcribed mRNA catabolic process, nonsense-mediated decay
GO:0019083~viral transcription
GO:0006364~rRNA processing
GO:0030855~epithelial cell differentiation
GO:0060071~Wnt signaling pathway, planar cell polarity pathway
GO:0090263~positive regulation of canonical Wnt signaling pathway
GO:0090090~negative regulation of canonical Wnt signaling pathway
GO:0000398~mRNA splicing, via spliceosome
GO:0006412~translation
GO:0042384~cilium assembly
GO:0060271~cilium morphogenesis
GO:0051301~cell division
GO:0006334~nucleosome assembly
GO:0006260~DNA replication
GO:0007062~sister chromatid cohesion
GO:0045944~positive regulation of transcription from RNA polymerase II promoter
GO:0000122~negative regulation of transcription from RNA polymerase II promoter
GO:0045893~positive regulation of transcription, DNA-templated
GO:0070830~bicellular tight junction assembly
GO:0007067~mitotic nuclear division
GO:0034080~CENP-A containing nucleosome assembly
GO:0007399~nervous system development
GO:0001764~neuron migration
GO:0030198~extracellular matrix organization
GO:0030199~collagen fibril organization
GO:0030574~collagen catabolic process
GO:0007155~cell adhesion
GO:0001501~skeletal system development
GO:0001649~osteoblast differentiation

Figure 3

A

Wong 2021

CA1

Dapi TTF1 SOX9 SOX2 Merged

Dapi TTF1 CFTR b-catenin Merged

NKX2-1GFP

Dapi TTF1 SOX9 SOX2 Merged

Dapi TTF1 CFTR b-catenin Merged

CA1

Dapi IgG IgG IgG Merged

B

CA1

2020

2015

NKX2-1GFP

2020

2015

C

Old

Agonist

CFTRInh

RFU Relative to Baseline (%)

Time (s)

DMSO

FSK

D

New

Agonist

CFTRInh

RFU Relative to Baseline (%)

Time (s)

DMSO

FSK

E

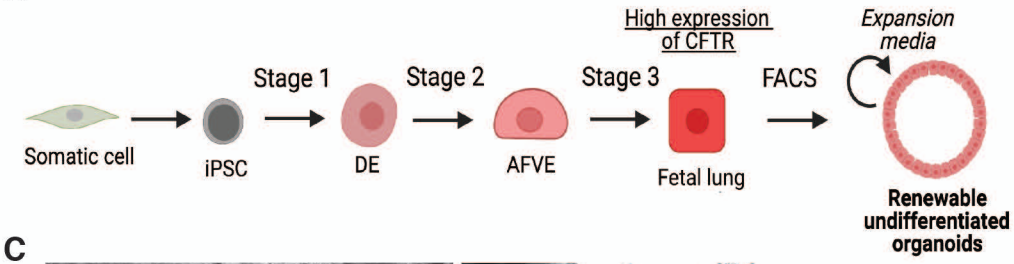
Maximum forskolin-induced response (%)

Old

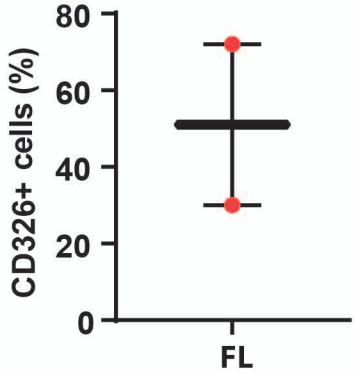
New

Figure 4

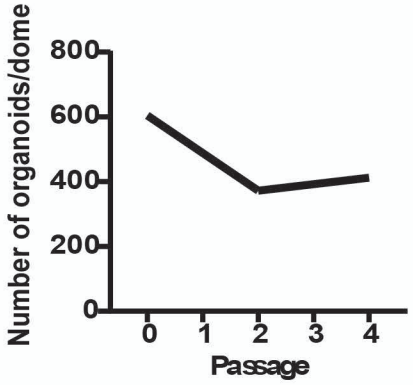
A



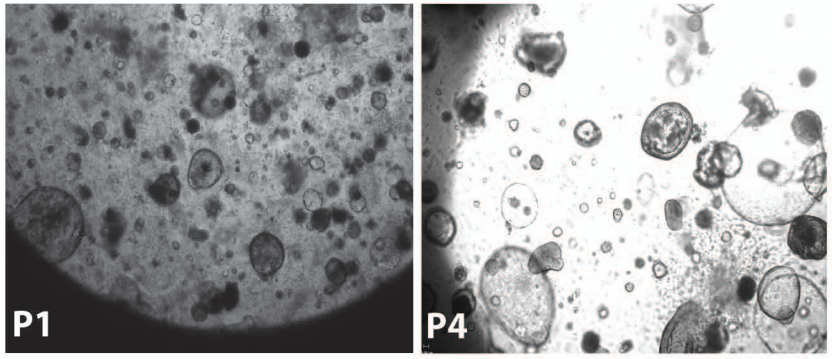
B



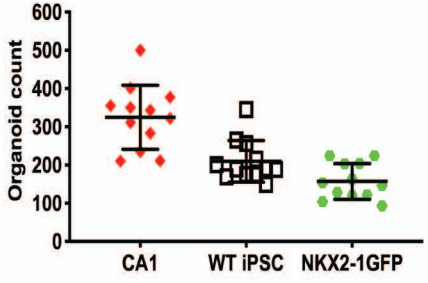
D



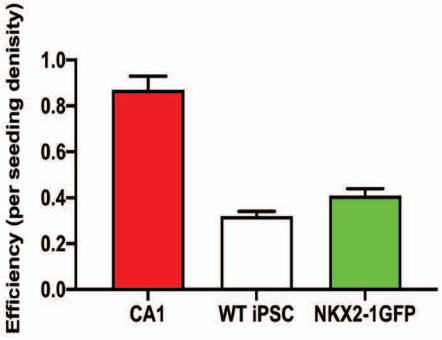
C



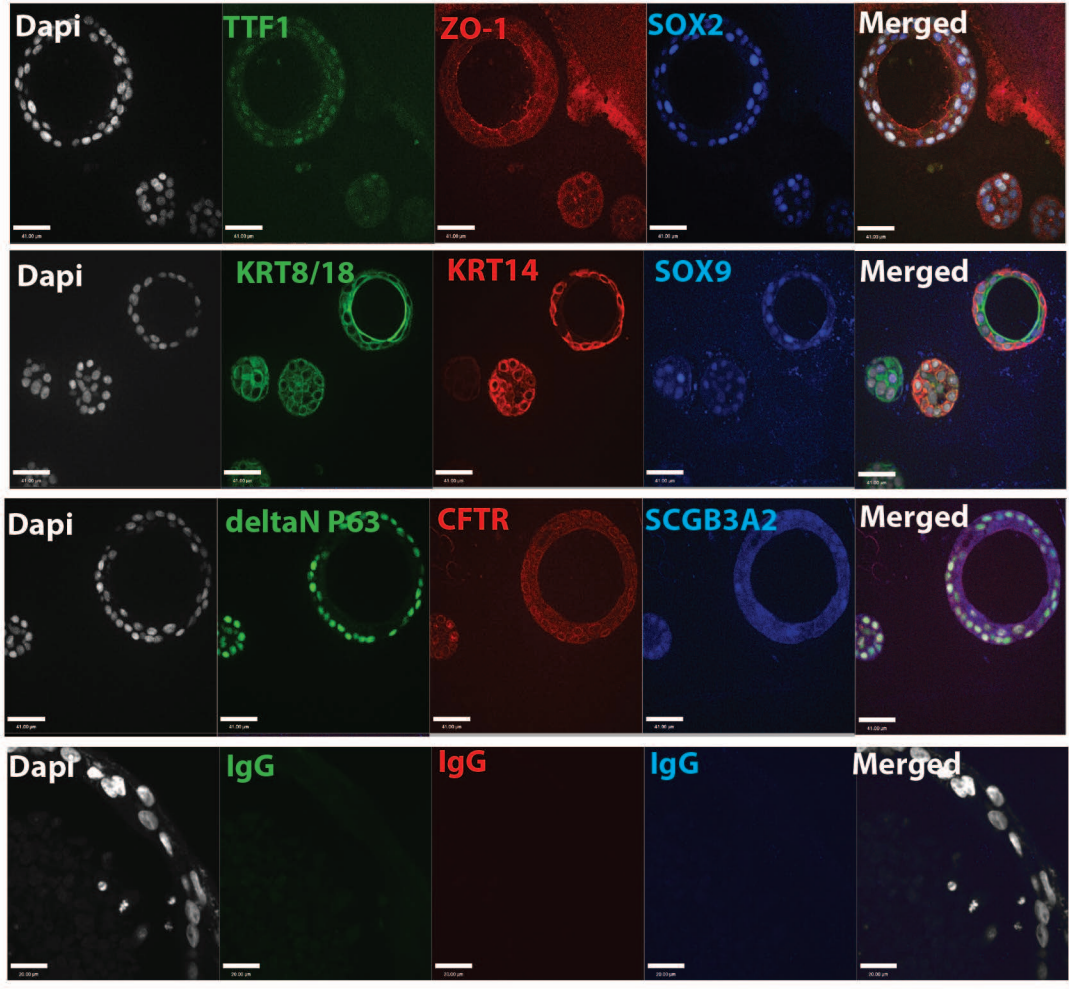
E



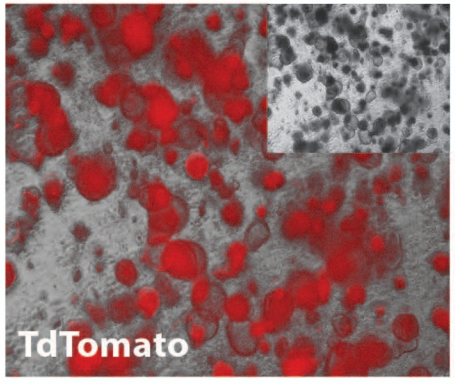
F



I



G



H

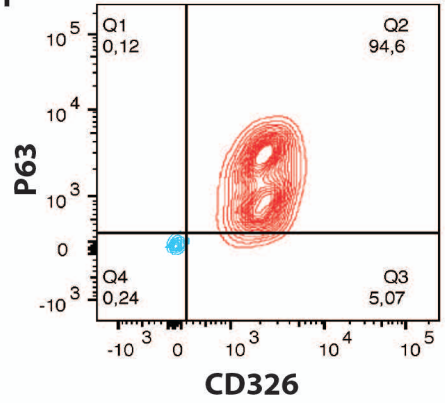


Figure 5

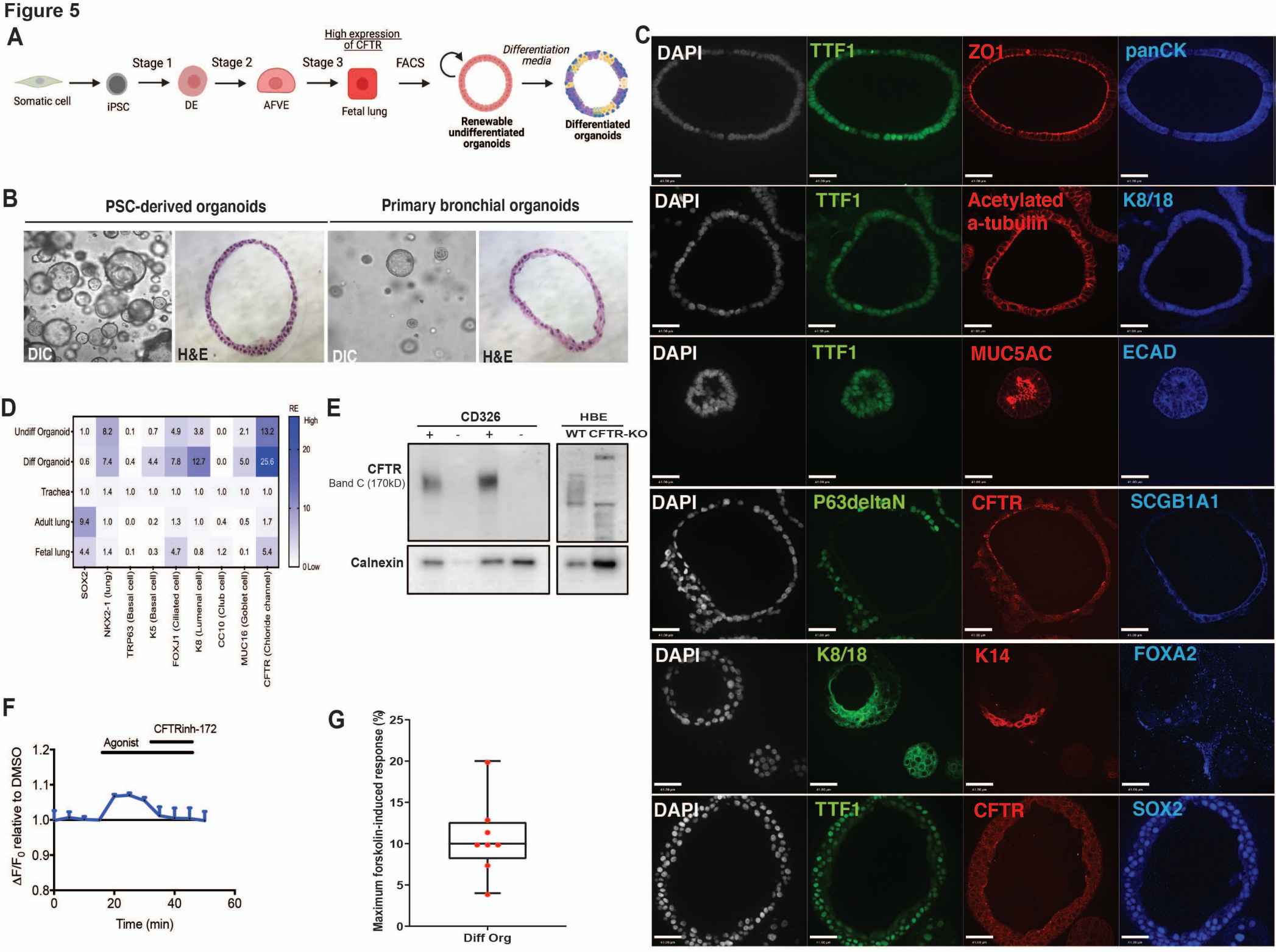

A
Somatic cell
iPSC
Stage 1
DE
Stage 2
AFVE
Stage 3
High expression of CFTR
Fetal lung
FACS
Renewable undifferentiated organoids
Differentiation media
Differentiated organoids
B
PSC-derived organoids
Primary bronchial organoids
DIC
H&E
DIC
H&E
C
DAPI
TTF1
ZO1
panCK
DAPI
TTF1
Acetylated a-tubulin
K8/18
DAPI
TTF1
MUC5AC
ECAD
DAPI
P63deltaN
CFTR
SCGB1A1
DAPI
K8/18
K14
FOXA2
DAPI
TTF1
CFTR
SOX2
D
RE
High
20
10
0 Low
Undiff Organoid 1.0 8.2 0.1 0.7 4.9 3.8 0.0 2.1 13.2
Diff Organoid 0.6 7.4 0.4 4.4 7.8 12.7 0.0 5.0 25.6
Trachea 1.0 1.4 1.0 1.0 1.0 1.0 1.0 1.0 1.0
Adult lung 9.4 1.0 0.0 0.2 1.3 1.0 0.4 0.5 1.7
Fetal lung 4.4 1.4 0.1 0.3 4.7 0.8 1.2 0.1 5.4
SOX2
NKX2-1 (lung)
TRP63 (Basal cell)
K5 (Basal cell)
FOXJ1 (Ciliated cell)
K8 (Lumenal cell)
CC10 (Club cell)
MUC16 (Goblet cell)
CFTR (Chloride channel)
E
CD326
+ - + -
HBE
WT CFTR-KO
CFTR
Band C (170kD)
Calnexin
F
CFTRinh-172
Agonist
ΔF/F0 relative to DMSO
1.2
1.1
1.0
0.9
0.8
0
20
40
60
Time (min)
G
Maximum forskolin-induced response (%)
25
20
15
10
5
0
Diff Org

Figure 6

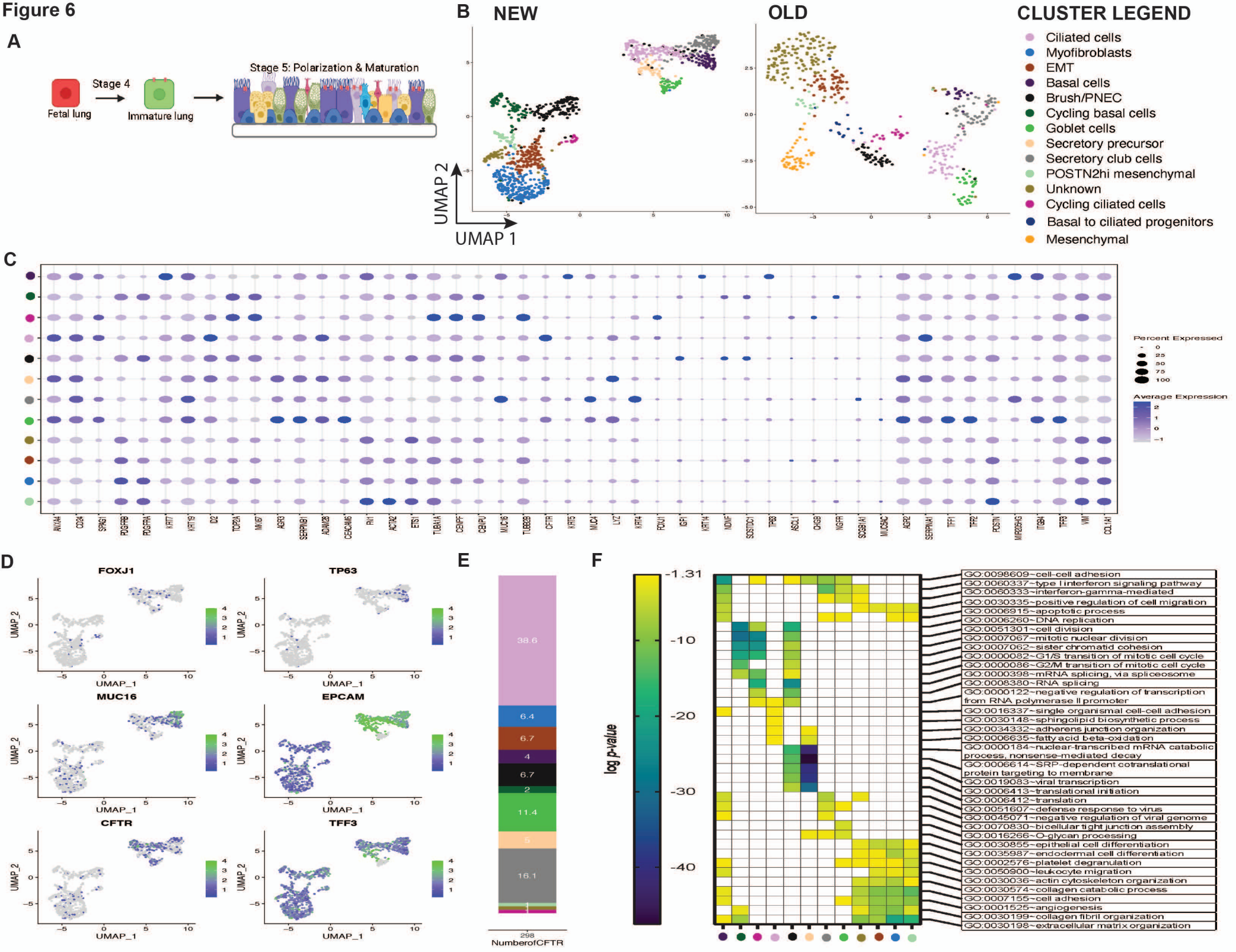
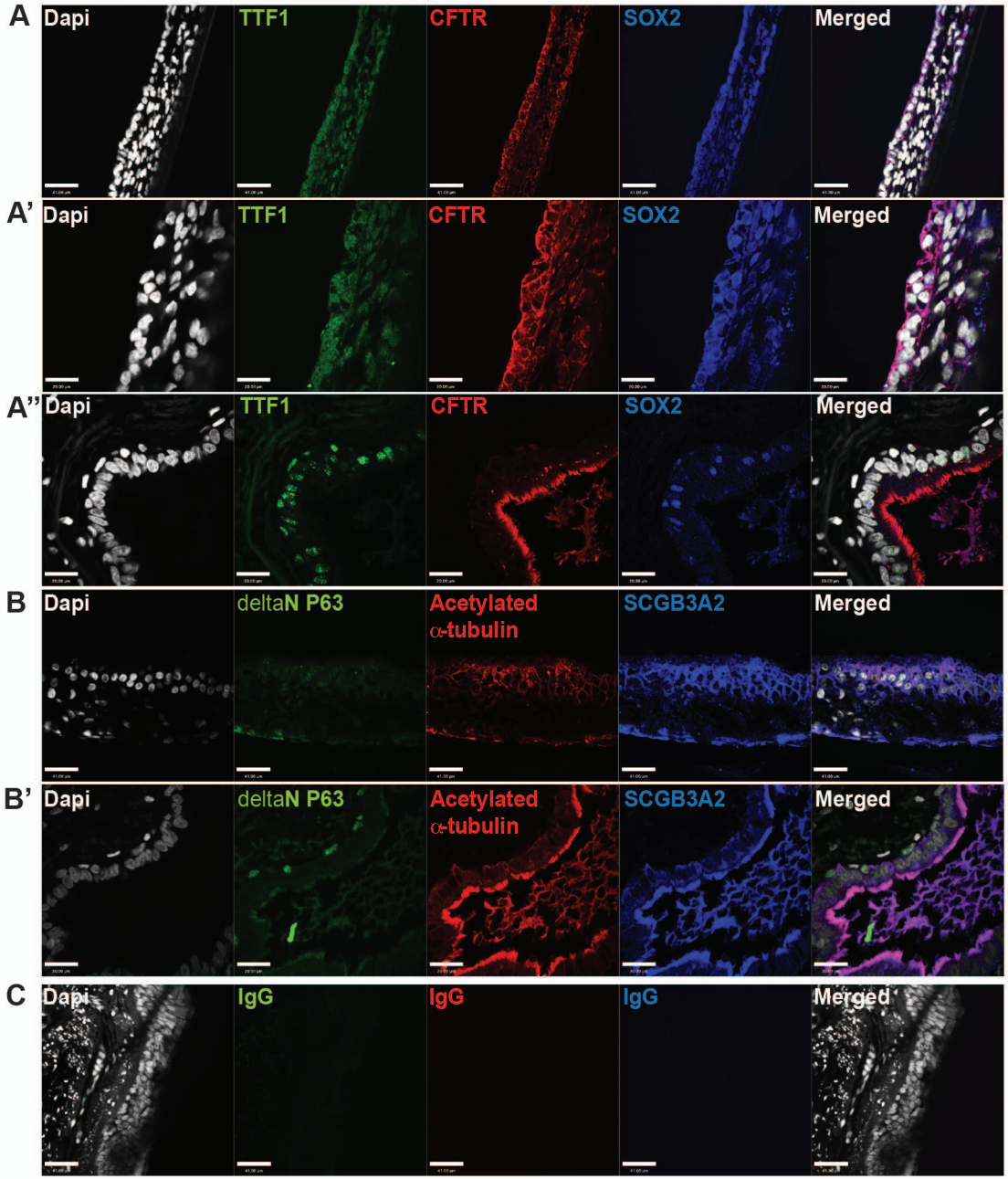
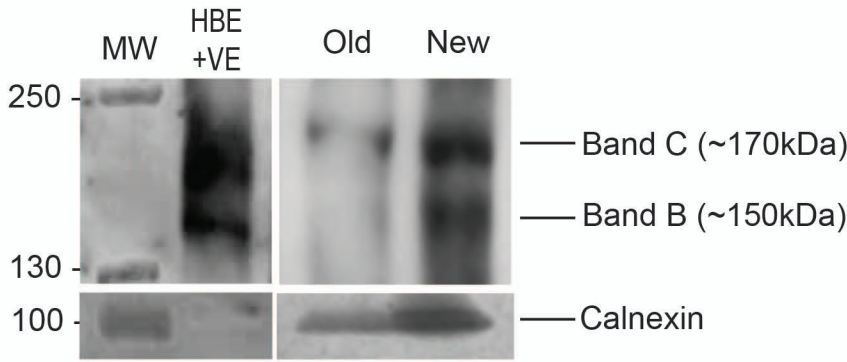
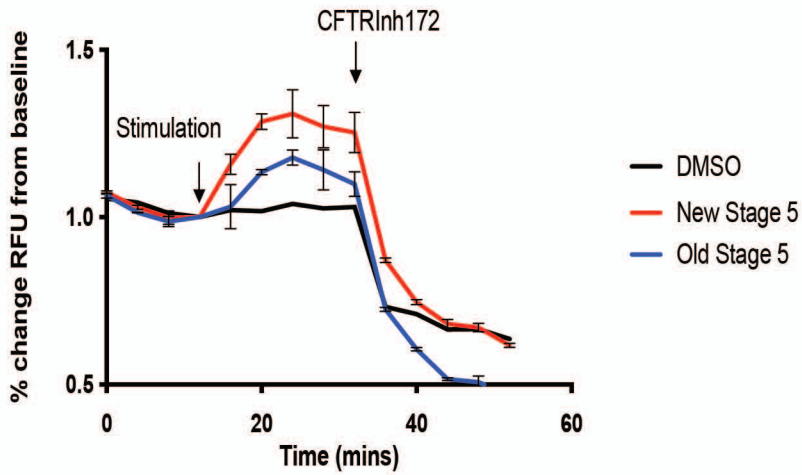

A
Fetal lung
Stage 4
Immature lung
Stage 5: Polarization & Maturation
B
NEW
OLD
UMAP 2
UMAP 1
CLUSTER LEGEND
Ciliated cells
Myofibroblasts
EMT
Basal cells
Brush/PNEC
Cycling basal cells
Goblet cells
Secretory precursor
Secretory club cells
POSTN2hi mesenchymal
Unknown
Cycling ciliated cells
Basal to ciliated progenitors
Mesenchymal
C
Percent Expressed
Average Expression
D
FOXJ1
TP63
MUC16
EPCAM
CFTR
TFF3
UMAP_1
UMAP_2
E
38.6
6.4
6.7
4
6.7
2
11.4
5
16.1
1
298
NumberofCFTR
F
-1.31
-10
-20
-30
-40
log p-value
GO:0098609~cell-cell adhesion
GO:0060337~type I interferon signaling pathway
GO:0060333~interferon-gamma-mediated
GO:0030335~positive regulation of cell migration
GO:0006915~apoptotic process
GO:0006260~DNA replication
GO:0051301~cell division
GO:0007067~mitotic nuclear division
GO:0007062~sister chromatid cohesion
GO:0000082~G1/S transition of mitotic cell cycle
GO:0000086~G2/M transition of mitotic cell cycle
GO:0000398~mRNA splicing, via spliceosome
GO:0008380~RNA splicing
GO:0000122~negative regulation of transcription from RNA polymerase II promoter
GO:0016337~single organismal cell-cell adhesion
GO:0030148~sphingolipid biosynthetic process
GO:0034332~adherens junction organization
GO:0006635~fatty acid beta-oxidation
GO:0000184~nuclear-transcribed mRNA catabolic process, nonsense-mediated decay
GO:0006614~SRP-dependent cotranslational protein targeting to membrane
GO:0019083~viral transcription
GO:0006413~translational initiation
GO:0006412~translation
GO:0051607~defense response to virus
GO:0045071~negative regulation of viral genome
GO:0070830~bicellular tight junction assembly
GO:0016266~O-glycan processing
GO:0030855~epithelial cell differentiation
GO:0035987~endodermal cell differentiation
GO:0002576~platelet degranulation
GO:0050900~leukocyte migration
GO:0030036~actin cytoskeleton organization
GO:0030574~collagen catabolic process
GO:0007155~cell adhesion
GO:0001525~angiogenesis
GO:0030199~collagen fibril organization
GO:0030198~extracellular matrix organization

Figure 7


A
Dapi
TTF1
CFTR
SOX2
Merged
A'
Dapi
TTF1
CFTR
SOX2
Merged
A''
Dapi
TTF1
CFTR
SOX2
Merged
B
Dapi
deltaN P63
Acetylated α-tubulin
SCGB3A2
Merged
B'
Dapi
deltaN P63
Acetylated α-tubulin
SCGB3A2
Merged
C
Dapi
IgG
IgG
IgG
Merged


D
MW
HBE +VE
Old
New
250
130
100
Band C (~170kDa)
Band B (~150kDa)
Calnexin


E
CFTRInh172
Stimulation
% change RFU from baseline
1.5
1.0
0.5
0
20
40
60
Time (mins)
DMSO
New Stage 5
Old Stage 5

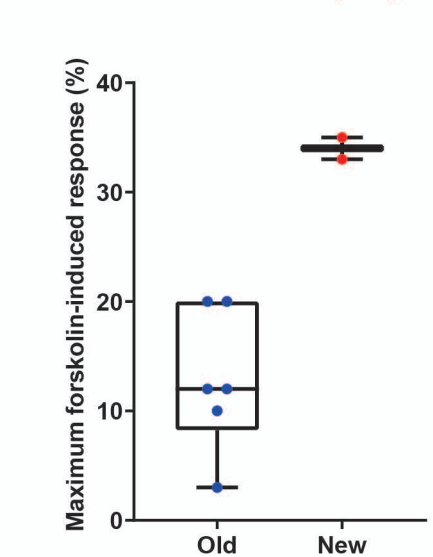
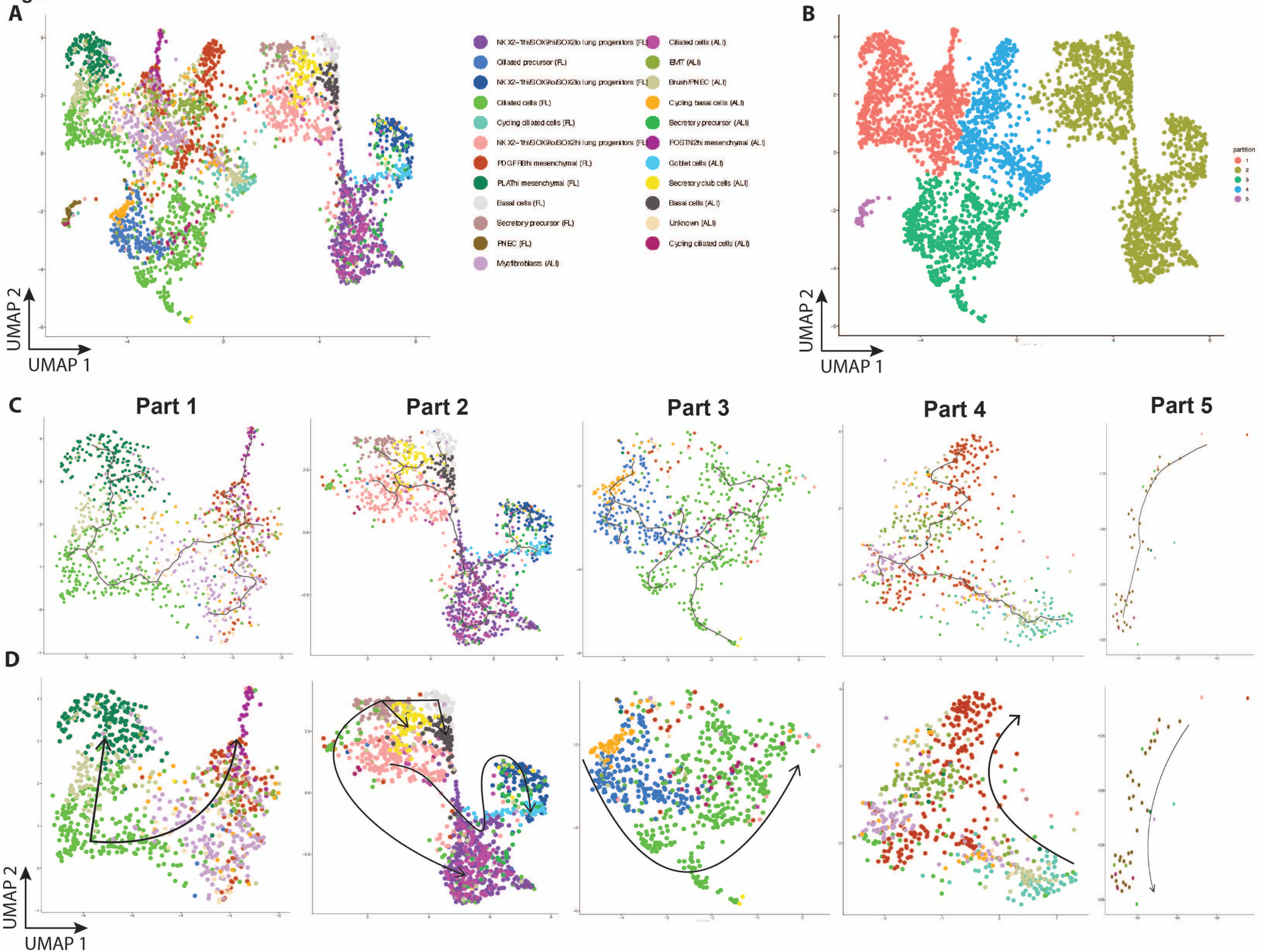

F
Maximum forskolin-induced response (%)
40
30
20
10
0
Old
New

Figure 8


A
NKX2-1hiSOX9hiSOX2lo lung progenitors (FL)
Ciliated precursor (FL)
NKX2-1hiSOX9loSOX2lo lung progenitors (FL)
Ciliated cells (FL)
Cycling ciliated cells (FL)
NKX2-1hiSOX9loSOX2hi lung progenitors (FL)
PDGFRBhi mesenchymal (FL)
PLAThi mesenchymal (FL)
Basal cells (FL)
Secretory precursor (FL)
PNEC (FL)
Myofibroblasts (ALI)
Ciliated cells (ALI)
EMT (ALI)
Brush/PNEC (ALI)
Cycling basal cells (ALI)
Secretory precursor (ALI)
POSTN2hi mesenchymal (ALI)
Goblet cells (ALI)
Secretory club cells (ALI)
Basal cells (ALI)
Unknown (ALI)
Cycling ciliated cells (ALI)
UMAP 2
UMAP 1
B
partition
1
2
3
4
5
UMAP 2
UMAP 1
C
Part 1
Part 2
Part 3
Part 4
Part 5
D
UMAP 2
UMAP 1